

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 15:14:42 ; Search time 8.70968 Seconds
(without alignments)
2938.192 Million cell updates/sec

Title: US-09-873-546-2

Perfect score: 33
Sequence: 1 gggggatccatgcgcgaacagagtaacattac 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications_NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.2	61.2	3249	10	US-09-925-301-457
2	19.4	58.8	249487	9	US-10-026-188-3
3	19.2	58.2	2081	10	US-09-925-301-233
4	19.2	58.2	5201	9	US-10-098-841-213
5	19.2	58.2	5404	9	US-09-954-531-130
6	19.2	58.2	5404	9	US-09-954-531-351
7	19.2	57.6	439	9	US-10-015-219-938
8	19.2	57.6	439	10	US-09-777-564-938
9	19.2	57.6	3266	9	US-09-989-545-9
10	18.2	55.2	1176	10	US-09-917-800A-1645
11	17.8	53.9	1479	10	US-09-883-797-3
12	17.8	53.9	6395	9	US-09-962-527-1
13	17.8	53.9	6425	9	US-09-962-527-3
14	17.8	53.9	6439	9	US-09-962-527-2
15	17.8	53.9	6446	9	US-09-962-527-5
16	17.8	53.9	6475	9	US-09-962-527-4
17	17.8	53.9	7685	9	US-09-949-317-22
18	17.8	53.9	7685	9	US-09-949-317-25
19	17.8	53.9	7685	9	US-09-949-316-22

20	17.8	53.9	7685	9	US-09-949-316-25	Sequence 25, Appl
21	17.8	53.9	7686	9	US-09-949-317-23	Sequence 23, Appl
22	17.8	53.9	7686	9	US-09-949-317-26	Sequence 26, Appl
23	17.8	53.9	7686	9	US-09-949-316-23	Sequence 23, Appl
24	17.8	53.9	7686	9	US-09-949-316-26	Sequence 26, Appl
25	17.8	53.9	7687	9	US-09-949-317-24	Sequence 24, Appl
26	17.8	53.9	7687	9	US-09-949-316-24	Sequence 24, Appl
27	17.8	53.9	7688	9	US-09-949-316-22	Sequence 22, Appl
28	17.8	53.9	7688	9	US-09-949-316-27	Sequence 27, Appl
29	17.8	53.9	10332	10	US-09-978-139-3	Sequence 3, Appl
30	17.4	52.7	350	10	US-09-867-701-695	Sequence 6095, Ap
31	17.4	52.7	552	10	US-09-867-701-3800	Sequence 3800, Ap
32	17.2	52.1	254	10	US-09-923-876-1950	Sequence 1950, Ap
33	17.2	52.1	468	10	US-09-770-444-240	Sequence 240, App
34	17.2	52.1	750	9	US-09-938-842A-21	Sequence 21, Appl
35	17.2	52.1	1592	9	US-09-954-531-1363	Sequence 1363, Ap
36	17.2	52.1	477	10	US-09-764-847-31	Sequence 31, Appl
37	17.2	52.1	2741	9	US-10-184-832-4	Sequence 4, Appl
38	17.2	52.1	6902	10	US-09-764-847-1019	Sequence 1019, Ap
39	17.2	52.1	22161	10	US-09-764-847-1020	Sequence 1020, Ap
40	16.8	50.9	270	10	US-09-878-574-7659	Sequence 7659, Ap
41	16.8	50.9	401	9	US-09-946-807-751	Sequence 751, App
42	16.8	50.9	401	9	US-09-946-807-752	Sequence 752, App
43	16.8	50.9	401	9	US-09-946-807-753	Sequence 753, App
44	16.8	50.9	401	10	US-09-795-668-751	Sequence 751, App
45	16.8	50.9	401	10	US-09-795-668-752	Sequence 752, App

ALIGNMENTS

```
RESULT 1
US-09-925-301-457
; Sequence 457, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 457
; LENGTH: 3249
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3234)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-457

Query Match          61.2%  Score 20.2;  DB 10;  Length 3249;
Best Local Similarity 75.8%;  Pred. No. 6.8;
Matches 25;  Conservative 0;  Mismatches 8;  Indels 0;  Gaps 0;

OY  1  GGGGATCCATGCCGGAACAGATGACGATTAC 33
    ||| ||||| ||||| ||||| ||| |||
Db   3001 GGCAGCTCCATGCTGGAACAGATGAGGAAATTC 3033

RESULT 2
US-10-026-188-3
; Sequence 3, Application US/10026188
; Patent No. US20020164645A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Zhang, Yifeng
```

```

; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Assays for Taste Receptor Cell Specific
; FILE REFERENCE: 02307E-114910US
; CURRENT APPLICATION NUMBER: US/10/026,188
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/259,379
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 249487
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse genomic region containing Itpc5
US-10-026-188-3

Query Match
Best Local Similarity 58.8%; Score 19.4; DB 9; Length 249487;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 2 GGGATCCATGCGCGAACAAGTAACGAT 30
Db 198467 GTGATCCATGCCAGACAGGAATGAT 198495

RESULT 3
US-09-925-301-233/c
; Sequence 233, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 233
; LENGTH: 2081
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-301-233

Query Match
Best Local Similarity 58.2%; Score 19.2; DB 10; Length 2081;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 GGGGATCCATGCGCGAACAAGTAACGATTA 32
Db 396 GGGGATCCATGCGCGAACAAGTAACGACTA 365

RESULT 4
US-10-098-841-213/c
; Sequence 213, Application US/10098841
; Publication No. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungding
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
```

```

; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Dimanac, Radoje T.
; TITLE OF INVENTION: No. US20020197679A1 Nucleic Acids and
; FILE REFERENCE: 784CIP2
; CURRENT APPLICATION NUMBER: US/10/098,841
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: PL_FL-genes Version 1.0
; SEQ ID NO 213
; LENGTH: 5201
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: CDS
; LOCATION: (847)..(3063)
US-10-098-841-213
```

```

Query Match
Best Local Similarity 58.2%; Score 19.2; DB 9; Length 5201;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 2 GGGATCCATGCGCGAACAAGTAACGATTAC 33
Db 4476 GGAGATCCAGCGAGTAAGTAAGTAAGTCTGAC 4445
```

```

RESULT 5
US-09-954-531-130/c
; Sequence 130, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/234,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 130
; LENGTH: 5404
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-130
```

```

Query Match
Best Local Similarity 58.2%; Score 19.2; DB 9; Length 5404;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 2 GGGATCCATGCGCGAACAAGTAACGATTAC 33
Db 3434 GGAGATCCAGCGAGTAAGTAAGTAAGTCTGAC 3403
```

```

RESULT 6
US-09-954-531-351/C
; Sequence 351, Application US/09954531
; Patent No. US20020165180A1
; GENERAL INFORMATION:
; APPLICANT: Weaver, Zoe
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cancer
; TITLE OF INVENTION: Gene Sets
; FILE REFERENCE: 689290-77
; CURRENT APPLICATION NUMBER: US/09/954,531
; CURRENT FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US/60/233,133
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,009
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,034
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,509
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: US/60/234,567
; PRIOR FILING DATE: 2000-09-22
; NUMBER OF SEQ ID NOS: 1392
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 351
; LENGTH: 5404
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-531-351

```

```

Query Match Similarity      58.2%; Score 19.2; DB 9; Length 5404;
Best Local Similarity       75.0%; Pred. No. 21;
Matches    24; Conservative    0; Mismatches    8; Indels    0; Gaps    0;

OY      2   GGGATCCATGCCGGAACAGACTAAGATTAC 33
          ||||| | | | | | | | | | | | |
Db      3434 GGAGATCCAGCGCAGTAAGTAACTGTGC 3403

RESULT 7
US-10-015-219-938/c
; Sequence 938, Application US/10015219
; Publication No. US20030008299A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.493CI
; CURRENT FILING DATE: US/10/015.219
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 938
; LENGTH: 439
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-015-219-938

Query Match          57.6%; Score 19; DB 9; Length 439;
Best Local Similarity 81.5%; Pred. No. 18;
Matches    22; Conservative    0; Mismatches    5; Indels    0; Gaps    0;

OY      3   GGGATCCATGCCGGAACAGACTAACGA 29
          ||||| | | | | | | | | | | | |
Db      97   GGGATCCATTCTGGAACAAGATGAAGA 71

RESULT 8
US-09-777-564-938/c
; Sequence 938, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
```

```

1  APPLICANT:  Algate, Paul A.
2  APPLICANT:  Mannion, Jane
3  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY
4  TITLE OF INVENTION:  AND DIAGNOSIS OF OVARIAN CANCER
5  FILE REFERENCE:  210121.493
6  CURRENT APPLICATION NUMBER:  US/09/777,564
7  CURRENT FILING DATE:  2001-02-05
8  NUMBER OF SEQ ID NOS:  1730
9  SOFTWARE:  FastSeq for Window Version 4.0
10 SEQ ID NO 938
11 LENGTH: 439
12 TYPE: DNA
13 ORGANISM: Homo sapiens
14 US-09-777-564-938

```

Query Match	57.6%	Score 19	DB 10	Length 439
Best Local Similarity	81.5%	Pred. NO. 18		
Matches 22	Conservative 0	Mismatches 5	Indels 0	Gaps 0
QY	3	GGGATCCATCCCGGGAACGAGTACGA	29	
DB	97	GGGATCCATTCGTGGGAACGAGTAAAGA	71	

```

RESULT 9
US-09-989-545-9/c
; Sequence 9, Application US/09989545
; Patent No. US20020164697A1
; GENERAL INFORMATION:
; APPLICANT: Lehar, Sophie
; APPLICANT: Manning, Stephen
; APPLICANT: Coyle, Anthony J.
; APPLICANT: Gutierrez-Ramos, Jose-Carlos
; TITLE OF INVENTION: No. US20020164697A1e1 Th2-Specific Molecules and Uses thereof
; FILE REFERENCE: 5800-10B
; CURRENT APPLICATION NUMBER: US/09/989,545
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 09/168,229
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: 09/258,670
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 3266
; TYPE: DNA
; ORGANISM: Mus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (40)..(642)
; NAME/KEY: unsure
; LOCATION: (1)..(3266)
; OTHER INFORMATION: n= a, g, c, or t
US-09-989-545-9

Query Match 57.6%; Score 19; DB 9; Length 3266;
Best Local Similarity 81.5%; Pred. No. 24;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 GGGGATCCATCCCGAGACAGACTAACG 28
||||||| | ||||||||| |
Db 2366 GGGGATCCTTCAGAGACAGACTAAAG 2340

RESULT 10
US-09-917-800A-1645/c
; Sequence 1645, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur

```


;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Howrey & Simon
;; STREET: 1299 Pennsylvania Avenue N.W.
;; CITY: Washington
;; STATE: DC
;; COUNTRY: USA
;; ZIP: 20004
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq for Windows Version 2.0
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/962,527
;; FILING DATE: 24-Sep-2001
;; CLASSIFICATION: <Unknown>
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/037,751
;; FILING DATE: 10-march-1998
;;
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Halluin, Albert P
;; REGISTRATION NUMBER: 25,277
;; REFERENCE/DOCKET NUMBER: 00801.0140.999
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650-463-8109
;; TELEFAX: 650-463-8400
;; TELEX: <Unknown>
;;
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 6425 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: unknown
;; MOLECULE TYPE: Genomic RNA
;; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-962-527-3
Query Match 53.9%; Score 17.8; DB 9; Length 6425;
Best Local Similarity 58.6%; Pred. No. 97;
Matches 17; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
QY 5 GATCCATGCCGAGACAGTACGATTAC 33
Db 3398 GGUCGAGCAGGACACAAUAGCAUUAUC 3426
RESULT 14
US-09-962-527-2
Sequence 2, Application US/09962527
Publication No. US20030049813A1
GENERAL INFORMATION:
APPLICANT: GARGER, STEPHEN
HOLTZ, R. BARRY
MCULLOCH, MICHAEL
TURNER, THOMAS
TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
FROM PLANT SOURCES
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/962,527

;; FILING DATE: 24-Sep-2001
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/037,751
;; FILING DATE: 10-march-1998
;;
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Halluin, Albert P
;; REGISTRATION NUMBER: 25,277
;; REFERENCE/DOCKET NUMBER: 00801.0140.999
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650-463-8109
;; TELEFAX: 650-463-8400
;; TELEX: <Unknown>
;;
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 6439 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: unknown
;; MOLECULE TYPE: Genomic RNA
;; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-962-527-2
Query Match 53.9%; Score 17.8; DB 9; Length 6439;
Best Local Similarity 58.6%; Pred. No. 97;
Matches 17; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
QY 5 GATCCATGCCGAGACAGTACGATTAC 33
Db 3398 GGUCGAGCAGGACACAAUAGCAUUAUC 3426
RESULT 15
US-09-962-527-5
Sequence 5, Application US/09962527
Publication No. US20030049813A1
GENERAL INFORMATION:
APPLICANT: GARGER, STEPHEN
HOLTZ, R. BARRY
MCULLOCH, MICHAEL
TURNER, THOMAS
TITLE OF INVENTION: A PROCESS FOR ISOLATING AND
PURIFYING VIRUSES SOLUBLE PROTEINS AND PEPTIDES
FROM PLANT SOURCES
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/962,527
FILING DATE: 24-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/037,751
FILING DATE: 10-march-1998
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 00801.0140.999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-463-8109
TELEFAX: 650-463-8400
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 6446 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: genomic RNA
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
us-09-962-527-5

Query Match 53.9%; Score 17.8; DB 9; Length 6446;
Best Local Similarity 58.6%; Pred. No. 97;
Matches 17; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 5 GATCCATGCCGACAGAGTAACGATTAC 33
| : | : | | | | | | : | | : | |
DB 3398 GGUCGAUCCAGACACACAUAAGCAUUDAC 3426

Search completed: March 23, 2003, 17:16:40
Job time : 81.7097 secs

GenCore version 5.1.4-ps.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 14:20:51 ; Search time 214.548 Seconds
(without alignments)
2491.051 Million cell updates/sec

Title: US-09-873-546-2

Perfect score: 33
Sequence: 1 gggggatccatgcgcgaacagagtaacgattac 33

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

otal number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	25	75.8	454	9	AA723009 zH30e05.s
2	25	75.8	470	9	A1497811 tB89f05.x
3	25	75.8	664	12	BG699685 602681538
4	22.6	68.5	466	10	BE019848 bb60e12.y
5	21.8	66.1	410	13	BI350717 fr37c06.y
6	21.8	66.1	467	12	BG892366 fp75d04.y

7	21.8	66.1	558	13	BM034832	BM034832 f36f03.y
8	21.8	66.1	573	13 <td>BM186599</td> <td>BM186599 fv75a03.y</td>	BM186599	BM186599 fv75a03.y
9	21.8	66.1	626	10 <td>AM134357</td> <td>AM134357 f18d11.y</td>	AM134357	AM134357 f18d11.y
10	21.8	66.1	628	13 <td>BM172422</td> <td>BM172422 imagegc.4</td>	BM172422	BM172422 imagegc.4
11	21.8	66.1	632	13 <td>BI475251</td> <td>BI475251 fg30d09.y</td>	BI475251	BI475251 fg30d09.y
12	21.8	66.1	645	13 <td>BI547693</td> <td>BI547693 603191841</td>	BI547693	BI547693 603191841
13	21.8	66.1	702	10 <td>AM305720</td> <td>AM305720 f163a05.y</td>	AM305720	AM305720 f163a05.y
14	21.8	66.1	709	12 <td>BG699614</td> <td>BG699614 602679344</td>	BG699614	BG699614 602679344
15	21.8	66.1	786	13 <td>BI552244</td> <td>BI552244 603195901</td>	BI552244	BI552244 603195901
16	21.8	66.1	801	12 <td>BG702913</td> <td>BG702913 602684858</td>	BG702913	BG702913 602684858
17	21.8	66.1	893	14 <td>BQ922385</td> <td>BQ922385 AGENCOURT</td>	BQ922385	BQ922385 AGENCOURT
18	21.8	66.1	1013	14 <td>BQ881423</td> <td>BQ881423 AGENCOURT</td>	BQ881423	BQ881423 AGENCOURT
19	21	63.6	441	13 <td>BM283669</td> <td>BM283669 k150f02.y</td>	BM283669	BM283669 k150f02.y
20	21	63.6	571	17 <td>AZ143597</td> <td>AZ143597 SP_0043.B</td>	AZ143597	AZ143597 SP_0043.B
21	21	63.6	597	13 <td>BI594830</td> <td>BI594830 As t92.44</td>	BI594830	BI594830 As t92.44
22	21	63.6	613	9 <td>BM665049</td> <td>BM665049 K191d01.y</td>	BM665049	BM665049 K191d01.y
23	20.2	61.2	174	9 <td>AA341050</td> <td>AA341050 EST46459</td>	AA341050	AA341050 EST46459
24	20.2	61.2	309	14 <td>F06359</td> <td>F06359 HSC15F101 n</td>	F06359	F06359 HSC15F101 n
25	20.2	61.2	312	10 <td>AM249313</td> <td>AM249313 2819320.5</td>	AM249313	AM249313 2819320.5
26	20.2	61.2	320	14 <td>F07725</td> <td>F07725 HSC2FG101 n</td>	F07725	F07725 HSC2FG101 n
27	20.2	61.2	336	9 <td>AA309505</td> <td>AA309505 EST180604</td>	AA309505	AA309505 EST180604
28	20.2	61.2	355	10 <td>BE394856</td> <td>BE394856 601310612</td>	BE394856	BE394856 601310612
29	20.2	61.2	359	14 <td>F06503</td> <td>F06503 HSC1AD101 n</td>	F06503	F06503 HSC1AD101 n
30	20.2	61.2	364	14 <td>F06857</td> <td>F06857 HSC1MH111 n</td>	F06857	F06857 HSC1MH111 n
31	20.2	61.2	372	9 <td>AA172156</td> <td>AA172156 2P24f11.f</td>	AA172156	AA172156 2P24f11.f
32	20.2	61.2	396	10 <td>AV684237</td> <td>AV684237 AV684237</td>	AV684237	AV684237 AV684237
33	20.2	61.2	408	10 <td>BE266662</td> <td>BE266662 601193986</td>	BE266662	BE266662 601193986
34	20.2	61.2	410	9 <td>A1400668</td> <td>A1400668 t989d04.x</td>	A1400668	A1400668 t989d04.x
35	20.2	61.2	415	10 <td>AV736284</td> <td>AV736284 AV736284</td>	AV736284	AV736284 AV736284
36	20.2	61.2	417	9 <td>AA312757</td> <td>AA312757 EST183405</td>	AA312757	AA312757 EST183405
37	20.2	61.2	426	14 <td>T62496</td> <td>T62496 YC03907.r1</td>	T62496	T62496 YC03907.r1
38	20.2	61.2	450	9 <td>AA232077</td> <td>AA232077 xz23q08.f</td>	AA232077	AA232077 xz23q08.f
39	20.2	61.2	462	12 <td>BF193212</td> <td>BF193212 244717.MA</td>	BF193212	BF193212 244717.MA
40	20.2	61.2	467	10 <td>AV715589</td> <td>AV715589 AV715589</td>	AV715589	AV715589 AV715589
41	20.2	61.2	470	14 <td>R15962</td> <td>R15962 ya47f08.r1</td>	R15962	R15962 ya47f08.r1
42	20.2	61.2	493	9 <td>AL711183</td> <td>AL711183 DRK2P686K</td>	AL711183	AL711183 DRK2P686K
43	20.2	61.2	515	14 <td>BM837087</td> <td>BM837087 K-ESTF0113</td>	BM837087	BM837087 K-ESTF0113
44	20.2	61.2	524	14 <td>BM846631</td> <td>BM846631 K-ESTF0125</td>	BM846631	BM846631 K-ESTF0125
45	20.2	61.2	536	9 <td>AL561640</td> <td>AL561640 AL561640</td>	AL561640	AL561640 AL561640

ALIGNMENTS

RESULT 1
AA723009 454 bp mRNA linear EST 02-JAN-1998
LOCUS zH30e05.s1 Soares_pinea1_gland_N3HPG Homo sapiens cDNA clone
DEFINITION IMAGE:413600 3', mRNA sequence.
ACCESSION AA723009
VERSION AA723009.1 GI:2740716
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 454)
AUTHORS Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Maria,W., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.lbnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amerisham
High quality sequence stop: 453.

[illegible]

TITLE	JOURNAL
COMMENT	
Query Match	Best Local Similarity 92.0%; Pred. No. 55;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
OY 9 CATGCCGACAGCACTAATCATTAC 33 Db 230 CATGCCAGACGACAGTAACGATTAC 254	
RESULT 6	
BG892366	467 bp mRNA linear EST 13-FEB-2002
LOCUS	fpt5d04.y1 zebrafish adult retina cDNA Danio rerio cDNA clone
DEFINITION	4787766.5 similar to TR:095057 095057 BC41195.1.; mRNA sequence.
ACCESSION	BG892366
VERSION	BG892366.1 GI:14286976
KEYWORDS	EST.
SOURCE	zebrafish.
ORGANISM	Danio rerio
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
AUTHORS	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
	; Cyprinidae; Danio.
	1 (bases 1 to 467)
	Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Maria,M., Eddy
	,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
	,K., Stepoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
	Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schuck,R., Ritter,E.,
	Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
	and Wilson,R.
	Washu zebrafish EST Project 1998
TITLE	JOURNAL
COMMENT	Contact: Stephen L. Johnson
	Washington University School of Medicine
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
	Tel: 314 286 1800
	Fax: 314 286 1810
	Email: zbrafish@wustl.edu
	cDNA Library Preparation: John Ngai, cDNA Library Arrayed by:
	Matthew Clark, DNA Sequencing by: Washington University Genome
	Sequencing Center Clone distribution: Genome Systems, St. Louis,
	Missouri (web address: www.genomesystems.com) (email contact:
	info@genomesystems.com) and Research Genetics, Huntsville, Alabama
	(web address: www.resgen.com) (email contact: info@resgen.com) and
	ResourceCenter@primat Datenbank, Berlin, Germany (web address:
	www.rtpd.de)
	Seq primer: T7
	High quality sequence stop: 288.
FEATURES	Location/Qualifiers
SOURCE	1..410
	/organism="Danio rerio"
	/db_xref="taxon:7955"
	/clone="4955530"
	/clone_1lb="zebrafish adult brain"
	/sex="mixed male and female"
	/tissue_type="brain"
	/dev_stage="adult"
	/lab_host="E. coli DH10B"
	/note="Vector: pZIRPlox; Site_1: NotI; site_2: SalI;
	Original library was constructed in lambdaZIRPlox. Mass
	excision of the cDNA library was performed to yield
	pZIRPlox plasmids. Insert check was done in original
	library."
BASE COUNT	116 a 98 c 101 g 92 t 3 others
ORIGIN	

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: zbrafish@wustl.wustl.edu

Library constructed by: Chandra Tucker and Gregory Niemi DNA Sequencing by: Washington University Genome Sequencing Center Clome distribution: RessourcenzentrumPrimatdatenbank, Berlin, Germany (web address: www.rzpd.de)

High quality sequence stop: 390.

FEATURES
Location/Qualifiers

1..467

/organism="Danio rerio"

/strain="wild-type"

/db_xref="taxon:7955"

/clone="4787766"

/clone_lib="zebrafish adult retina cDNA"

/sex="mixed"

/dev_stage="1-2 years"

/lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"

/note="Vector: lambda ZAP II (pluscript SK-); Site 1:

ECORI; Site 2: SalI; This zebrafish library was

constructed by Dr. Susan E. Brockhoff (email:

sbrockhoff@u.washington.edu) RZPD library number: 760"

BASE COUNT 125 a 124 c 118 g 99 t 1 others

ORIGIN

Query Match

Best Local Similarity 92.0%; Score 21.8; DB 12; Length 467;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 CATGCCGACAGACGATACGATTAC 33

||||| |||||||

Db 188 CATGCCGACAGACGATACGATTAC 212

RESULT 7

BM034832

LOCUS

558 bp mRNA linear EST 26-JUL-2002

fu36f03.y3 zebrafish adult brain Danio rerio cDNA clone 5331604 5'

similar to TR:095057 095057 BC41195_1.; mRNA sequence.

ACCESSION

BM034832

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

COMMENT

TITLE

JOURNAL

Other-ESTs: fu36f03.x1

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@wustl.wustl.edu

CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clome distribution: Genome Systems, St. Louis,

Missouri (web address: www.genomesystems.com) (email contact:

info@genomesystems.com) and Research Genetics, Huntsville, Alabama

(web address: www.resgen.com) (email contact: info@resgen.com) and

RessourcenzentrumPrimatdatenbank, Berlin, Germany (web address:

www.rzpd.de)

Seq primer: T7

High quality sequence stop: 437.

Location/Qualifiers

1..558

/organism="Danio rerio"

/db_xref="taxon:7955"

/clone="5331604"

/clone_lib="zebrafish adult brain"

/sex="mixed male and female"

/tissue_type="brain"

/dev_stage="adult"

/lab_host="E. coli DH10B"

/note="Vector: pZiPlox; Site 1: NotI; Site 2: SalI;

Original library was constructed in lambdaZiPlox. Mass

excision of the cDNA library was performed to yield

pZiPlox plasmids. Insert check was done in original

library."

BASE COUNT 150 a 153 c 137 g 118 t

ORIGIN

Query Match

Best Local Similarity 92.0%; Score 21.8; DB 13; Length 558;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 CATGCCGACAGACGATACGATTAC 33

||||| |||||||

Db 228 CATGCCGACAGACGATACGATTAC 252

RESULT 8

BM186599

LOCUS

573 bp mRNA linear EST 26-JUL-2002

fu75a03.y1 zebrafish adult brain Danio rerio cDNA clone 5414284 5'

similar to TR:095057 095057 BC41195_1.; mRNA sequence.

ACCESSION

BM186599

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

COMMENT

TITLE

JOURNAL

Other-ESTs: fu75a03.x1

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@wustl.wustl.edu

CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:

Matthew Clark. DNA Sequencing by: Washington University Genome

Sequencing Center Clome distribution: Genome Systems, St. Louis,

Missouri (web address: www.genomesystems.com) (email contact:

info@genomesystems.com) and Research Genetics, Huntsville, Alabama

(web address: www.resgen.com) (email contact: info@resgen.com) and

RessourcenzentrumPrimatdatenbank, Berlin, Germany (web address:

www.rzpd.de)

Seq primer: T7

High quality sequence stop: 449.

Location/Qualifiers

1..573

/organism="Danio rerio"

/db_xref="taxon:7955"

/clone="5414284"

/clone_lib="zebrafish adult brain"

/sex="mixed male and female"

```

/tissue_type="brain"
/dev_stage="adult"
/lab_host="E. coli DH10B"
/Note="Vector: pZIRPlox; Site_1: NotI; Site_2: SalI;
original library was constructed in lambdaZIRPlox. Mass
excision of the cDNA library was performed to yield
pZIRPlox plasmids. Insert check was done in original
library."
BASE COUNT      151 a      157 c      144 g      121 t
ORIGIN
Query Match      66.1%; Score 21.8; DB 13; Length 573;
Best Local Similarity 92.0%; Pred. No. 58;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 9 CATGCCGACAGAGTAACGATTAC 33
||||| || ||||| ||||| |||||
Db 228 CATGCCGACAGAGTAACGATTAC 252

RESULT 9
AM134357      626 bp      mRNA      linear      EST 27-OCT-1999
LOCUS
DEFINITION    f118d11.y1 Sugano Kawakami zebrafish DNA Danio rerio cDNA clone
ACCESSION    2601525.5 similar to TR:095057 095057 BC41195.1.; mRNA sequence.
VERSION      AM134357
KEYWORDS      AM134357.1 GI:6135964
SOURCE        zebrafish.
ORGANISM      Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 626)
Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
Martin,J., Page,D., Steptoe,M., Underwood,K., Theising,B., Riltter
,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
Washu Zebrafish EST Project 1999
Unpublished (1999)
Other_ESTS: f118d11.x1
Contact: S.L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center
Seq primer: T3 ET from Amersham
High quality sequence stop: 509.
Location/Qualifiers
1. 626
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone_id="2601525"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/Note="Vector: pME18S-FL3; Site_1: DraIII (CACCGCTG);
Site_2: DraIII (CACCAGCTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TCGTGGCTTCTG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACGTGCTG, 3' site
CACCAGCTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for

```

```

sequencing: 5' end primer CTCTGCTCTTAAAGCTCG and 3' end
primer CGACCTGACCTGACGACA.
BASE COUNT      168 a      166 c      167 g      123 t      2 others
ORIGIN
Query Match      66.1%; Score 21.8; DB 10; Length 626;
Best Local Similarity 92.0%; Pred. No. 59;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 9 CATGCCGACAGAGTAACGATTAC 33
||||| || ||||| ||||| |||||
Db 209 CATGCCGACAGAGTAACGATTAC 233

RESULT 10
BM172422      628 bp      mRNA      linear      EST 04-DEC-2001
LOCUS
DEFINITION    imagec4_2001/smu47bdf41.x1 NIH_MGC_95 Homo sapiens cDNA clone
IMAGE:4816673 5', mRNA sequence.
VERSION      BM172422
KEYWORDS      BM172422.1 GI:17311985
SOURCE        EST.
ORGANISM      human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 628)
Kale,P.I., Harsch,T.J., Folta,P.A., Nelson,D.O., Sanders,C.G. and
Prange,C.K.
The I.M.A.G.E. Consortium quality control effort: clone
resequencing for verification
Unpublished (2001)
Contact: Prange CK
The I.M.A.G.E. Consortium
Lawrence Livermore National Laboratory
Livermore, CA, USA
Email: help@image.llnl.gov
This read has been produced as part of the I.M.A.G.E. Consortium
quality control effort. High quality sequence is defined as having
100 or more base pairs with a phred quality value of 20 or greater,
where a sliding window of 4 base pairs with a phred quality value
of 15 or greater marks the beginning and end of the sequence. For
information on obtaining this clone, please contact
info@image.llnl.gov.
Plate: ILAM10715 row: 1 column: 18
Seq primer: -21m13
High quality sequence stop: 628.
Location/Qualifiers
1. 628
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="4816673"
/clone_id="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/Note="Organ: brain; Vector: pBluescriptPR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (Gcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTNN-3',
size selected for average insert size 2.5 kb and
normalized to ROP 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NIHRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      139 a      175 c      197 g      116 t
ORIGIN
Query Match      66.1%; Score 21.8; DB 13; Length 628;
Best Local Similarity 92.0%; Pred. No. 59;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 9 CATGCCGACAGAGTAACGATTAC 33
||||| || ||||| ||||| |||||

```

DB 129 CATGCTGATCAGAGTACGATTAC 153

RESULT 11
BI475251

LOCUS
DEFINITION f1q30a09.y3 zebrafish adult brain Danio rerio cDNA clone 4832920 5' similar to TR:095057 095057 BC41195_1. ; mRNA sequence.

ACCESSION
VERSION BI475251
KEYWORDS BI475251.1 GI:15306381
SOURCE EST.
ORGANISM zebrafish.
Danio rerio

REFERENCE
AUTHORS Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

TITLE Washu zebrafish EST Project 1998

JOURNAL Unpublished (1998)

COMMENT Other_ESTS: f1q30a09.x3
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
CDNA library preparation: John Ngai. CDNA library Arrayed by: Matthew Clark. DNA sequencing by: Washington University Genome Sequencing Center distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and Resourcenzentrum Primatendatenbank, Berlin, Germany (web address: www.rzp.de)
Seq primer: T7
High quality sequence stop: 495.

FEATURES
source
location/Qualifiers
1..632
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_id="4832920"
/clone_lib="zebrafish adult brain"
/sex="mixed male and female"
/tissue_type="brain"
/dev_stage="adult"
/lab_host="E. coli DH10B"
/note="Vector: pZiPlox; Site_1: NotI; Site_2: SalI; Original library was constructed in lambdaZiPlox. Mass excision of the cDNA library was performed to yield pZiPlox plasmids. Insert check was done in original library."

BASE COUNT 172 a 169 c 164 g 127 t

ORIGIN

Query Match 66.1%; Score 21.8; DB 13; Length 632;
Best Local Similarity 92.0%; Pred. No. 59;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 CATGCCGAGACAGTACGATTAC 33
||||| |||||||
DB 227 CATGCCAGACGAGTACGATTAC 251

RESULT 12
BI547693

LOCUS
DEFINITION B1547693 645 bp mRNA linear EST 05-SEP-2001 603191841P1 NIH_MGC_95 Homo sapiens CDNA IMAGE:5262995 5',

ACCESSION
VERSION BI547693
KEYWORDS BI547693.1 GI:15435005
SOURCE EST.
ORGANISM human.
Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NIH-MGC http://mhc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shireki Toshiyuki and Piero Carninci (RIKEN)

COMMENT DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLM11663 row: a column: 12
High quality sequence stop: 634.

FEATURES
source
location/Qualifiers
1..645
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="IMAGE:5262995"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 161 a 166 c 205 g 113 t

ORIGIN

Query Match 66.1%; Score 21.8; DB 13; Length 645;
Best Local Similarity 92.0%; Pred. No. 59;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 CATGCCGAGACAGTACGATTAC 33
||||| |||||||
DB 74 CATGCTGAGCAGAGTACGATTAC 98

RESULT 13
AM305720

LOCUS
DEFINITION f1q30a05.y1 zebrafish adult brain Danio rerio cDNA 5' similar to TR:095057 095057 BC41195_1. ; mRNA sequence.

ACCESSION
VERSION AM305720
KEYWORDS AM305720.1 GI:6718073
SOURCE EST.
ORGANISM zebrafish.
Danio rerio

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
1 (bases 1 to 702)
Clark, M., Johnson, S. L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and Wilson, R.

TITLE Washu Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu

CDNA Library Preparation: John Ngai. CDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
Resourcenet@uniprimar Datenbank, Berlin, Germany (web address:
www.rzpd.de) '77
Seq primer: '77
High quality sequence stop: 486.
Location/Qualifiers

FEATURES

source

1. .702
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish adult brain"
/sex="mixed male and female"
/tissue_type="brain"
/dev_stage="adult"
/lab_host="E. coli DH10B"
/note="Vector: pZIRPlox; Site.1: NotI; Site.2: SalI;
original library was constructed in lambdaZIRPlox. Mass
excision of the cDNA library was performed to yield
pZIRPlox plasmids. Insert check was done in original
library."
BASE COUNT 194 a 189 c 181 g 137 t 1 others
ORIGIN

Query Match 66.1%; Score 21.8; DB 10; Length 702;
Best Local Similarity 92.0%; Pred. No. 60;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 CATGCCGACAGAGTAACGATTAC 33
||||| |||||||
Db 219 CATGCCAGACAGAGTAACGATTAC 243

RESULT 14 709 bp mRNA linear EST 07-MAY-2001
LOCUS BG699614
DEFINITION 60267934F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:481962 5',
mRNA sequence.
ACCESSION BG699614
VERSION BG699614.1 GI:13968106
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 709)
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM10703 row: h column: 11
High quality sequence stop: 701.
Location/Qualifiers

FEATURES

source

1. .709
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:481962"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVA-3',
size-selected for average insert size 2.5 kb and
normalized to R0T 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 168 a 191 c 226 g 124 t
ORIGIN

Query Match 66.1%; Score 21.8; DB 12; Length 709;
Best Local Similarity 92.0%; Pred. No. 60;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 9 CATGCCGACAGAGTAACGATTAC 33
||||| |||||||
Db 150 CATGCCAGACAGAGTAACGATTAC 174

RESULT 15 786 bp mRNA linear EST 05-SEP-2001
BI552244
LOCUS 603195901F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5275297 5',
mRNA sequence.
ACCESSION BI552244
VERSION BI552244.1 GI:15439556
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 786)
TITLE NIH-MGC http://mgi.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: LLM1695 row: b column: 02
High quality sequence stop: 589.
Location/Qualifiers

FEATURES

source

1. .786
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5275297"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site.1: BamHI; Site.2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTVA-3',
size-selected for average insert size 2.5 kb and
normalized to R0T 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
FEATURES

BASE COUNT 182 a 228 c 238 g 138 t
 ORIGIN

Query Match 66.1%; Score 21.8; DB 13; Length 786;
 Best Local Similarity 92.0%; Pred. No. 61;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 9 CATGCCGACAGAGTACGATTAC 33
 ||||| ||||| ||||| ||||| |||||
 Db 73 CATGCCGACAGAGTACGATTAC 97

Search completed: March 23, 2003, 17:12:20
 Job time : 220.548 secs

REFERENCE 1
 AUTHORS Chalup, M.S., Altus, C.M., Lincoln, S.E., Dufour, G.E. and Jackson, S.
 TITLE Molecules for disease detection and treatment
 JOURNAL Patent: WO 0240715-A 200 23-MAY-2002;
 INCYTE GENOMICS INC (US)

FEATURES
 source 1. .3061
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID NO: LI:332161.1:2000SEP08"

BASE COUNT 606 a 840 c 927 g 688 t
 ORIGIN

Query Match 75.8%; Score 25; DB 6; Length 3061;
 Best Local Similarity 84.8%; Pred. No. 1.1;
 Matches 28; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 GGGGATCCATGCCGGACAGTAACGATTAC 33
 b 42 GCGGGAGAGATGCCGGACAGTAACGATTAC 74

RESULT 5
 LOCUS BC030660 3391 bp mRNA linear PRI 21-MAY-2002
 DEFINITION Homo sapiens, similar to R1g protein, clone MGC:33391
 IMAGE:481437, mRNA, complete cds.

ACCESSION BC030660.1 GI:21040534
 VERSION BC030660.1
 KEYWORDS MGC.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 TITLE 1 (bases 1 to 3391)
 JOURNAL Strausberg, R.
 Direct Submission
 Submitted (20-MAY-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
 CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
 Toshiyuki and Ptero Carninci (RIKEN)
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
 contact: amadan@systemsbiology.org
 Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
 Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whitting
 Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAX Plate: 46 Row: e Column: 8
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Genomescan gene
 prediction.

FEATURES
 source 1. .3391
 /organism="Homo sapiens"
 /db_xref="locusID:148252"
 /db_xref="taxon:9606"
 /clone="MGC:33391 IMAGE:481437"
 /tissue_type="Brain, hippocampus"
 /clone_lib="NIH MGC_95"
 /lab_host="DH10B"
 /note="Vector: pBluescript"
 138. .734
 /codon_start=1

/product="similar to R1g protein"
 /protein_id="AAH30660.1"
 /db_xref="GI:21040534"
 /translation="MEQSDNDRVAVVEGAGVGSSTVLFYKGERPDYPTIDTY
 ROVISGDKSVCTLOITDTGSHQPRAMQDLSTKSHAFILVSVYSKQSELELCITYK
 LTVQIRGSVEDIPVMLVGNKCDGTQREAFQRAQAAQDMKCAFMTSAKNYNVKEL
 FOELLTLEFRRMNSLINDGRSGKGRKRDVRVAKCTLM"

BASE COUNT 650 a 996 c 1016 g 729 t

Query Match 75.8%; Score 25; DB 9; Length 3391;
 Best Local Similarity 84.8%; Pred. No. 1.1;
 Matches 28; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 GGGGATCCATGCCGGACAGTAACGATTAC 33
 Db 129 GCGGGAGAGATGCCGGACAGTAACGATTAC 161

RESULT 6
 LOCUS AC006538/c 177540 bp DNA linear PRI 07-FEB-1999
 DEFINITION Homo sapiens chromosome 19, BAC 41195 (CIT-B-31c16), complete
 sequence.

ACCESSION AC006538
 VERSION AC006538.1 GI:4235145
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 TITLE 1 (bases 1 to 177540)
 JOURNAL Lamerdin, J.E., McCready, P.M., Skowronski, E., Wiseman, V.,
 Burkhardt-Schultz, K., Gordon, L., Dias, J., Ramirez, M., Stiliwagen, S.,
 Phan, H., Velasco, N., Do, L., Regala, M., Terry, A., Barnes, J.,
 Dangnanan, L., Erler, A., Christensen, M., Georgescu, A., Avila, J.,
 Liu, S., Altix, C., Andreise, T., Trankheim, M., Amico-Keller, G.,
 Coefficient, J., Duarte, S., Lucas, S., Bruce, R., Thomas, P., Quan, G.,
 Krommiller, B., Arellano, A., Sanders, C., Ow, D., Nolan, M., Trong, S.,
 Kobayashi, A., Olsen, A.S. and Carrino, A.V.

Sequence analysis of a 3.5 kb contig in human 19p13.3 containing a
 serine protease gene cluster

TITLE Unpublished
 JOURNAL 2 (bases 1 to 177540)
 REFERENCE Lamerdin, J.E.
 AUTHORS Lamerdin, J.E.
 TITLE Direct Submission
 JOURNAL Submitted (07-FEB-1999) Joint Genome Institute, Lawrence Livermore
 National Laboratory, 7000 East Ave., Livermore, CA 94551, USA

Map and sequence oriented from p telomere to centromere. BC41195
 (CIT-B-31c16) is currently separated from cosmid R32203 (AC006275)
 to the left by a sequence gap of approximately 6 kb, and overlaps
 BAC 102889 (CIT-B-191n6, AC006130) to the right from bases 175,545
 to 177,540. Additional map and sequence information are available
 at: <http://www.bio.lnl.gov/db/rrp/genome/genome.html>.

FEATURES
 source 1. .177540
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /map="19p13.3 between CDC34 and D19S342"
 /clone="CIT-B-31c16"
 /clone_lib="Cal tech CIT-B BAC library"
 /note="LNL clone name: BC41195"
 Location/Qualifiers
 repeat_region complement(1..144)
 repeat_region /rpt_family="LMC/D"
 repeat_region 283..381
 repeat_region /rpt_family="AluJo"
 repeat_region complement(399..437)
 repeat_region /rpt_family="(CAAA)n"
 repeat_region complement(438..738)
 repeat_region /rpt_family="AluSx"
 repeat_region complement(868..931)
 repeat_region /rpt_family="L1MB7"

misc_feature 1175. .1365
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: good, score: 57.000"
repeat_region 1608. .1904
/rpt_family="AluSg"
complement(2133. .2231)
repeat_region /rpt_family="MER81"
2739. .2935
/rpt_family="AluJo"
repeat_region 2946. .3247
/rpt_family="AluSx"
repeat_region 3256. .3384
/rpt_family="L1MC3"
repeat_region 3389. .3604
/rpt_family="AluJo"
repeat_region 3631. .3716
/rpt_family="L1MD3"
complement(4087. .4242)
repeat_region /rpt_family="AluJo"
complement(4294. .4436)
repeat_region /rpt_family="AluJo"
complement(4598. .4887)
repeat_region /rpt_family="AluSg"
complement(4906. .5373)
repeat_region /rpt_family="L1MB7"
complement(5382. .5669)
repeat_region /rpt_family="AluSx"
complement(6425. .6609)
repeat_region /rpt_family="(GGCA)n"
complement(6610. .6765)
repeat_region /rpt_family="(GAAA)n"
complement(6766. .7065)
repeat_region /rpt_family="AluSx"
7093. .7392
repeat_region /rpt_family="AluSx"
7423. .7511
repeat_region /rpt_family="L1MA7"
7539. .7839
repeat_region /rpt_family="AluSx"
7852. .7938
repeat_region /rpt_family="L1MA7"
8001. .8312
repeat_region /rpt_family="AluJo"
complement(8771. .8897)
repeat_region /rpt_family="MER20"
complement(9532. .9722)
repeat_region /rpt_family="MER20"
complement(9970. .10143)
repeat_region /rpt_family="MER20"
10144. .10405
repeat_region /rpt_family="AluJo"
10421. .10630
repeat_region /rpt_family="L1M3"
10631. .10756
repeat_region /rpt_family="AluSg"
10757. .11058
repeat_region /rpt_family="AluY"
11068. .11243
repeat_region /rpt_family="AluSg"
11247. .11520
repeat_region /rpt_family="L1M3"
11522. .11816
repeat_region /rpt_family="AluSg"
complement(11827. .12033)
repeat_region /rpt_family="MER58A"
12255. .12555
repeat_region /rpt_family="AluJo"
complement(12562. .12617)
repeat_region /rpt_family="(TA)n"
12769. .12894
repeat_region /rpt_family="FLAM_A"
complement(12901. .12951)
repeat_region /rpt_family="L1MC3"

repeat_region 12960. .13013
/rpt_family="LTR13"
repeat_region 13013. .13408
/rpt_family="LTR13"
complement(13409. .13457)
repeat_region /rpt_family="L1MD3"
13651. .13908
repeat_region /rpt_family="L1MB8"
13906. .14116
repeat_region /rpt_family="L1MC2"
complement(14173. .14468)
repeat_region /rpt_family="AluY"
14497. .14811
repeat_region /rpt_family="AluSx"
14812. .15011
repeat_region /rpt_family="(GGCA)n"
15689. .16007
repeat_region /rpt_family="AluJo"
16008. .16128
repeat_region /rpt_family="MLT1F"
16130. .16430
repeat_region /rpt_family="AluSp"
16432. .16563
repeat_region /rpt_family="(GAAA)n"
16573. .16891
repeat_region /rpt_family="AluSx"
16892. .16930
repeat_region /rpt_family="(GAAA)n"
16989. .17073
repeat_region /rpt_family="MLT1F"
17178. .17479
repeat_region /rpt_family="AluSx"
17481. .18206
repeat_region /rpt_family="MER21B"
18240. .18329
repeat_region /rpt_family="MLT1F"
18380. .18748
repeat_region /rpt_family="MSTD"
complement(19030. .19229)
repeat_region /rpt_family="MER20"
complement(19579. .19692)
misc_feature /note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: good, score: 53.000"
repeat_region complement(19953. .20160)
repeat_region /rpt_family="MER20"
complement(20388. .20585)
repeat_region /rpt_family="MER20"
20846. .20994
repeat_region /rpt_family="MIR"
complement(21068. .21332)
repeat_region /rpt_family="AluSx"
complement(21372. .21671)
repeat_region /rpt_family="AluSg"
complement(21681. .21811)
repeat_region /rpt_family="FLAM_C"
21837. .22056
repeat_region /rpt_family="MER58A"
22240. .22539
repeat_region /rpt_family="AluSx"
22697. .22831
repeat_region /rpt_family="AluJo"
22833. .23132
repeat_region /rpt_family="AluSx"
23136. .23412
repeat_region /rpt_family="L1MC3"
complement(24002. .24172)
repeat_region /rpt_family="MER20"
complement(24191. .24313)
repeat_region /rpt_family="MIR"
24410. .24707
repeat_region /rpt_family="AluSg"
complement(24739. .24926)
repeat_region /rpt_family="MER58A"

repeat_region 24928..25229
/rpt_family="Alusx"

Query Match 75.8%; Score 25; DB 9; Length 177540;
Best Local Similarity 84.8%; Pred. No. 0.99;
Matches 28; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GGGGATCATCGCGAAGACAGATACGATTAC 33
|||||
Db 73015 GCGGGGAAGATCGCGAAGACAGATACGATTAC 72983

RESULT 7
AY056037
LOCUS Homo sapiens Rlg protein mRNA, complete cds. PRI 24-JUL-2002
ACCESSION AY056037
VERSION AY056037.1 GI:16555333
KEYWORDS Homo sapiens.
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 597)
AUTHORS Ellis,C.A., Vos,M.D., Howell,H., Vallecorsa,T., Fults,D.W. and Clark,G.J.
TITLE Rlg is a novel Ras-related protein and potential neural tumor suppressor
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 9876-9881 (2002)
JOURNAL 12107278
PUBMED 2 (bases 1 to 597)
REFERENCE Castro,A.F. and Quilliam,L.A.
AUTHORS Identification of Rlg, a novel Ras family member sharing homology with NOG2 and Rap
Unpublished
3 (bases 1 to 597)
REFERENCE Castro,A.F. and Quilliam,L.A.
AUTHORS Direct Submission
TITLE Submitted (13-SEP-2001) Biochemistry and Molecular Biology, Indiana University School of Medicine, 635 Barnhill Drive, MS-4053, Indianapolis, IN 46202, USA
JOURNAL Location/Qualifiers
FEATURES
source
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13.3"
/tissue_type="brain; anaplastic oligodendroglioma"
/note="derived from dbEST AI497811"
1..597
/note="Ras family GTPase"
/codon_start=1
/product="Rlg protein"
/protein_id="AL23715.1"
/db_xref="GI:16555334"
/translation="MPEOSNDYRVYVVGAGVGKSSLVLPFGVGTPEPTYIPTEDRYROYISCDKSVCTLOITDPTGSHOPAPARISISKGAFILVFSYTSKQSLPELGPYIKLIVQIKSSVEDIPVNLGNKCDTQREVDYREAQVAQEMKCAEMTSAKMNVKELFOELLTETRRNMSLNIDGKRSKQKRTDVKKCTLM"

BASE COUNT 134 a 178 c 195 g 90 t
ORIGIN

Query Match 72.7%; Score 24; DB 9; Length 597;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 ATGCCGGAACAGATACGATTAC 33
|||||
Db 1 ATGCCGGAACAGATACGATTAC 24

RESULT 8
AY059641

LOCUS AY059641 597 bp mRNA linear PRI 29-OCT-2001
DEFINITION Homo sapiens small GTP-binding tumor suppressor 1 mRNA, complete cds.
ACCESSION AY059641
VERSION AY059641.1 GI:16508175
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 597)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE Molecular cloning of GBR1, a novel gene encoding a small GTP-binding tumor suppressor
Unpublished
2 (bases 1 to 597)
REFERENCE Gong,L.
AUTHORS Direct Submission
TITLE Submitted (12-OCT-2001) Dept. of Cardiology, U.T. MD. Anderson Cancer Center, 1515 Holcombe Blvd-449, Houston, TX 77030, USA
JOURNAL Location/Qualifiers
FEATURES
source
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
1..597
/note="GBR1"
/codon_start=1
/product="small GTP-binding tumor suppressor 1"
/protein_id="AA17968.1"
/db_xref="GI:16508176"
/translation="MPEOSNDYRVYVVGAGVGKSSLVLPFGVGTPEPTYIPTEDRYROYISCDKSVCTLOITDPTGSHOPAPARISISKGAFILVFSYTSKQSLPELGPYIKLIVQIKSSVEDIPVNLGNKCDTQREVDYREAQVAQEMKCAEMTSAKMNVKELFOELLTETRRNMSLNIDGKRSKQKRTDVKKCTLM"

BASE COUNT 134 a 178 c 195 g 90 t
ORIGIN

Query Match 72.7%; Score 24; DB 9; Length 597;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 ATGCCGGAACAGATACGATTAC 33
|||||
Db 1 ATGCCGGAACAGATACGATTAC 24

RESULT 9
BC008065
LOCUS Homo sapiens, similar to CG8500 gene product, clone MGC:2364 IMAGE:3029902, mRNA, complete cds.
ACCESSION BC008065
VERSION BC008065.1 GI:14165524
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 1875)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
TITLE Direct Submission
JOURNAL Submitted (21-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
COMMENT Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DRP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology

http://www.systemsbio.org
 contact: amadan@systemsbiology.org
 Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
 Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
 through the T.M.A.G.E. Consortium/BLM: at: <http://limage.llnl.gov>
 Series: INAL Plate: 3 Row: p Column: 6.
 Location/Qualifiers

FEATURES

source

1..1875
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MGC:2364 IMAGE:3029902"
 /tissue_type="Ovary, adenocarcinoma"
 /clone_1lb="NH_MGC_9"
 /lab_host="DH10B-R"
 /note="Vector: pOTB7"
 80..679
 /codon_start=1
 /product="Similar to CG8500 gene product"
 /protein_id="AAH08065.1"
 /db_xref="GI:14165525"
 /translation="MEQSDNDYRAVAFGAGVGKSSLVLRFGYGFRESYIPTVEDY
 ROVISCDSICTLOITDTGSHQPPAMORLSTSKGAPLTVTSRSLSEKPIYE
 QICEIKGDESPIMLVGNKCDSPRSRQSEARALARTMKCAMETSAKLNHYKE
 LFQELNLEKRTVSLQIDGKSKQKREKLKGCVM"

CDS

BASE COUNT 518 a 438 c 472 g 447 t
 ORIGIN

Query Match 66.1%; Score 21.8; DB 9; Length 1875;
 Best Local Similarity 92.0%; Pred. No. 36;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

.QY 9 CATGCCGAGACAGATTAC 33
 ||||| || ||||| ||||| |||||
 Db 79 CATGCTGAGACAGATTAC 103

.RESULT 10
 AB076889 2505 bp mRNA linear PRI 25-JUN-2002
 LOCUS
 DEFINITION Homo sapiens mRNA for D1-Ras2, complete cds.
 ACCESSION AB076889
 VERSION AB076889.1 GI:21624249
 KEYWORDS
 SOURCE Homo sapiens adult brain cDNA to mRNA, clone_1lb:lambda ZAPII human brain cDNA.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS Kontani,K., Tada,M., Ogawa,T., Okai,T. and Katada,T.
 TITLE D1-Ras: A Distinct Subgroup of Ras-family GTPases with Unique
 Biochemical Properties
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2505)
 AUTHORS Kontani,K., Ogawa,T., Okai,T. and Katada,T.
 TITLE Direct Submission
 JOURNAL Submitted (24-DEC-2001) Toshiaki Katada, University of Tokyo,
 Department of Physiological Chemistry, Graduate School of
 Pharmaceutical Sciences, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033,
 Japan (E-mail:katadaemol.f.u-tokyo.ac.jp, Tel:81-3-5841-4750,
 Fax:81-3-5841-4751)

FEATURES

source

1..2505
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /Chromosome="9"
 /map="9q21.33-22.33"
 /tissue_type="Brain"
 /clone_1lb="lambda ZAPII human brain cDNA"
 /dev_stage="adult"
 1..2505
 gene

CDS

/gene="D1-ras2"
 31..630
 /gene="D1-ras2"
 /note="member of the Ras family
 small gtp-binding protein"
 /codon_start=1
 /product="D1-Ras2"
 /protein_id="BAC01116.1"
 /db_xref="GI:21624250"
 /translation="MEQSDNDYRAVAFGAGVGKSSLVLRFGYGFRESYIPTVEDY
 ROVISCDSICTLOITDTGSHQPPAMORLSTSKGAPLTVTSRSLSEKPIYE
 QICEIKGDESPIMLVGNKCDSPRSRQSEARALARTMKCAMETSAKLNHYKE
 LFQELNLEKRTVSLQIDGKSKQKREKLKGCVM"

BASE COUNT 622 a 611 c 630 g 642 t
 ORIGIN

Query Match 66.1%; Score 21.8; DB 9; Length 2505;
 Best Local Similarity 92.0%; Pred. No. 36;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

.QY 9 CATGCCGAGACAGATTAC 33
 ||||| || ||||| ||||| |||||
 Db 30 CATGCTGAGACAGATTAC 54

.RESULT 11
 AX477382 4167 bp DNA linear PAT 12-AUG-2002
 LOCUS
 DEFINITION Sequence 36 from Patent WO0231152.
 ACCESSION AX477382
 VERSION AX477382.1 GI:22216612
 KEYWORDS
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS Baughn,M.R., Ding,L., Elliott,V.S., Gandhi,A.R., Gietzen,K.J.,
 Griffin,J.A., Gururajan,R., Hatalla,A.J., Kearney,L., Khan,F.A.,
 Lal,P., Lee,E.A., Lu,D.A., Lu,Y., Nguyen,D.B., Arvizu,C.,
 Ramkumar,J., Tang,Y.T., Thangavelu,K., Thornton,M., Walla,N.K.,
 Warren,B.A., Xu,Y., Yao,M.G. and Yue,H.
 TITLE Intracellular signaling molecules
 JOURNAL Patent: WO 0231152-A 36 18-APR-2002;
 Incyte Genomics, Inc. (US)

FEATURES

source

1..4167
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 490576CB1"
 BASE COUNT 1098 a 911 c 960 g 1198 t
 ORIGIN

Query Match 66.1%; Score 21.8; DB 6; Length 4167;
 Best Local Similarity 92.0%; Pred. No. 35;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

.QY 9 CATGCCGAGACAGATTAC 33
 ||||| || ||||| ||||| |||||
 Db 160 CATGCTGAGACAGATTAC 184

FEATURES

LOCUS

AC100691 68052 bp DNA linear HTG 22-NOV-2001
 DEFINITION Mus musculus clone RP23-168121, LOW-PASS SEQUENCE SAMPLING.
 ACCESSION AC100691
 VERSION AC100691.1 GI:17048057
 KEYWORDS HTG; HTGS; PHASE0.
 SOURCE Mus musculus.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.


```

REFERENCE      1 (bases 1 to 68052)
AUTHORS        Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE          Mus musculus, clone RP23-168I21
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 68052)
AUTHORS        Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
                Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhaltier, B.,
                Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
                Chopel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A.,
                Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S.,
                Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gargyna, S.,
                Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
                Haags, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
                Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K.,
                Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
                Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,
                McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
                Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
                Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
                Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
                Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
                Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
                Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
                Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
                Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
                Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
                Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
                Direct Submission
                Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
                Research, 320 Charles Street, Cambridge, MA 02141, USA
                Smit, A.F.A. & Green, P. (1996-1997)
                http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L15958
Center clone name: 168_T_21

-----
* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 713 812: contig of 712 bp in length
* 813 1530: contig of 718 bp in length
* 1531 1630: gap of 100 bp
* 1631 2351: contig of 721 bp in length
* 2352 2451: gap of 100 bp
* 2452 3179: contig of 728 bp in length
* 3180 3279: gap of 100 bp
* 3280 3980: contig of 701 bp in length
* 3981 4080: gap of 100 bp
* 4081 4819: contig of 739 bp in length
* 4820 4919: gap of 100 bp
* 4920 5652: contig of 733 bp in length
* 5653 5752: gap of 100 bp
* 5753 6455: contig of 703 bp in length
* 6456 6555: gap of 100 bp
* 6556 7280: contig of 735 bp in length
* 7291 7390: gap of 100 bp
* 7391 8113: contig of 723 bp in length
* 8114 8213: gap of 100 bp

8214 8941: contig of 728 bp in length
8942 9041: gap of 100 bp
9042 9758: contig of 717 bp in length
9759 9858: gap of 100 bp
9859 10578: contig of 720 bp in length
10579 10678: gap of 100 bp
10679 11404: contig of 726 bp in length
11405 11504: gap of 100 bp
11505 12227: contig of 723 bp in length
12228 12327: gap of 100 bp
12328 13051: contig of 724 bp in length
13052 13151: gap of 100 bp
13152 13873: contig of 722 bp in length
13874 13973: gap of 100 bp
13974 14702: contig of 729 bp in length
14703 14802: gap of 100 bp
14803 15530: contig of 728 bp in length
15531 15630: gap of 100 bp
15631 16358: contig of 728 bp in length
16359 16458: gap of 100 bp
16459 17174: contig of 716 bp in length
17175 17274: gap of 100 bp
17275 17983: contig of 709 bp in length
17984 18083: gap of 100 bp
18084 18803: contig of 720 bp in length
18804 18903: gap of 100 bp
18904 19629: contig of 726 bp in length
19630 19729: gap of 100 bp
19730 20445: contig of 716 bp in length
20446 20545: gap of 100 bp
20546 21269: contig of 724 bp in length
21270 21369: gap of 100 bp
21370 22094: contig of 725 bp in length
22095 22194: gap of 100 bp
22195 22914: contig of 720 bp in length
22915 23014: gap of 100 bp
23015 23751: contig of 737 bp in length
23752 23851: gap of 100 bp
23852 24593: contig of 742 bp in length
24594 24693: gap of 100 bp
24694 25425: contig of 732 bp in length
25426 25525: gap of 100 bp
25526 26260: contig of 735 bp in length
26261 26360: gap of 100 bp
26361 27070: contig of 710 bp in length
27071 27170: gap of 100 bp
27171 27902: contig of 732 bp in length
27903 28002: gap of 100 bp
28003 28722: contig of 720 bp in length
28723 28822: gap of 100 bp
28823 29514: contig of 692 bp in length
29515 29614: gap of 100 bp
29615 30337: contig of 723 bp in length
30338 30437: gap of 100 bp
30438 31142: contig of 705 bp in length
31143 31242: gap of 100 bp
31243 31970: contig of 728 bp in length
31971 32070: gap of 100 bp
32071 32794: contig of 724 bp in length
32795 32894: gap of 100 bp
32895 33629: contig of 735 bp in length
33630 33729: gap of 100 bp
33730 34453: contig of 724 bp in length
34454 34553: gap of 100 bp
34554 35284: contig of 731 bp in length
35285 35384: gap of 100 bp
35385 36107: contig of 723 bp in length
36108 36207: gap of 100 bp
36208 36931: contig of 724 bp in length
36932 37031: gap of 100 bp
37032 37758: contig of 727 bp in length
37759 37858: gap of 100 bp
37859 38559: contig of 701 bp in length

```

```

* 38560 38659: gap of 100 bp
* 38660 39381: contig of 722 bp in length
* 39382 39481: gap of 100 bp
* 39482 40200: contig of 719 bp in length
* 40201 40300: gap of 100 bp
* 40301 41013: contig of 719 bp in length
* 41020 41119: gap of 100 bp
* 41120 41849: contig of 730 bp in length
* 41850 41949: gap of 100 bp
* 41950 42683: contig of 734 bp in length
* 42684 42783: gap of 100 bp
* 42784 43510: contig of 727 bp in length
* 43511 43610: gap of 100 bp
* 43611 44330: contig of 720 bp in length
* 44331 44430: gap of 100 bp
* 44431 45116: contig of 686 bp in length
* 45117 45216: gap of 100 bp
* 45217 45945: contig of 729 bp in length
* 45946 46045: gap of 100 bp
* 46046 46756: contig of 711 bp in length
* 46757 46856: gap of 100 bp
* 46857 47587: contig of 731 bp in length
* 47588 48410: contig of 723 bp in length
* 48411 48510: gap of 100 bp
* 48511 48236: contig of 726 bp in length
* 49237 49336: gap of 100 bp
* 49337 50016: contig of 680 bp in length
* 50017 50116: gap of 100 bp
* 50117 50814: contig of 698 bp in length
* 50815 50914: gap of 100 bp
* 50915 51639: contig of 725 bp in length
* 51640 51739: gap of 100 bp
* 51740 52463: contig of 724 bp in length
* 52464 52563: gap of 100 bp
* 52564 53280: contig of 717 bp in length
* 53281 53380: gap of 100 bp
* 53381 54093: contig of 713 bp in length
* 54094 54193: gap of 100 bp
* 54194 54915: contig of 722 bp in length
* 54916 55015: gap of 100 bp
* 55016 55720: contig of 705 bp in length
* 55721 55820: gap of 100 bp
* 55821 55534: contig of 714 bp in length
* 55535 56634: gap of 100 bp

```

```

Query Match      66.1%; Score 21.8; DB 2; Length 68052;
Best Local Similarity 92.0%; Pred. No. 32;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 9 CATGCCGACAGTACGATTAC 33
Db 4532 CATGCCGACAGTACGATTAC 4556

```

```

RESULT 13
LOCUS AL553619/c 87845 bp DNA linear PRI 13-NOV-2001
DEFINITION Human DNA sequence from clone Rp11-16P8 on chromosome 10, complete
sequence.
ACCESSION AL553619
VERSION AL553619.7 GI:16944093
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 87845)
AUTHORS Brown, J.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

```

COMMENT

On Nov 15, 2001 this sequence version replaced gi:15591326. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chr10>
 Rp11-16P8 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.choi.org/bacpac/home.htm>

VECTOR: pBAC3.6
 IMPORTANT: This sequence is not the entire insert of clone Rp11-16P8. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone Rp11-16P8 is at 1 in this sequence. The true left end of clone Rp11-551P18 is at 85846 in this sequence. The true right end of clone Rp11-422P15 is at 44164 in this sequence.

FEATURES

```

source      1..87845
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="10"
            /clone="Rp11-16P8"
            /clone_11b="RPCI-11.1"
BASE COUNT  26908 a 19283 c 18854 g 22800 t
ORIGIN

```

```

Query Match      66.1%; Score 21.8; DB 9; Length 87845;
Best Local Similarity 78.8%; Pred. No. 32;
Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 1 GGGGATCCATCCCGGACAGTACGATTAC 33
Db 22892 GGGGATCCATCCCGGACAGTACGATTAC 22924

```

```

RESULT 14
LOCUS AL553619/c 143299 bp DNA linear PRI 10-OCT-2001
DEFINITION Human DNA sequence from clone Rp11-563G12 on chromosome
9q21.33-22.33, complete sequence.
ACCESSION AL553619
VERSION AL553619.9 GI:16073639
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 143299)
AUTHORS Laird, G.
TITLE Direct Submission
JOURNAL Submitted (10-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 11, 2001 this sequence version replaced gi:15028633.
During sequence assembly data is compared from overlapping clones.

```

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Ch9> RP11-563G12 is from the library RPECI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/dacpac/home.htm> VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-563G12 it may be shorter because we sequence overlapping sections only once, except for a short overlap. The true right end of clone RP11-563G12 is at 143299 in this sequence. The true left end of clone RP11-61N16 is at 51116 in this sequence. The true right end of clone RP11-555F9 is at 2000 in this sequence.

FEATURES

source

1.143299
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="q21.33-q22.33"
/clone="RP11-563G12"
/clone_lib="RPECI-11.2"
4147..4230
misc-feature
/note="Single clone region. Assembly confirmed by restriction digest data."

BASE COUNT 43622 a 29112 c 29198 g 41367 t
ORIGIN

Query Match

Best Local Similarity 92.0%; Score 21.8; DB 9; Length 143299;

Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

9 CATGCCGACAGATACGATTAC 33
||||| || |||||||||
Db 55609 CATGCTGACAGATACGATTAC 55585

RESULT 15
AC021583/c 143857 bp DNA linear PRI 29-JUN-2002
LOCUS AC021583 Homo sapiens chromosome 9, clone RP11-1026H23, complete sequence.
DEFINITION AC021583
AC021583.11 GI:21629417
VERSION
KEYWORDS
SOURCE
ORGANISM
human
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,

TITLE

JOURNAL

REFERENCE

AUTHORS

Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heeford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karats,A., Klein,J., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J., Menus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K., Pierre,N., Pisenl,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (16-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 143857)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campiano,A., Chang,I., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karats,A., Kells,C., Labocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Roselli,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zalnoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

REFERENCE

AUTHORS

Submitted (21-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 143857)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karats,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Menus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zalnoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (29-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 29, 2002 this sequence version replaced: gi:21536044.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center Project name: L5759
 Center clone name: 1026_H23

SP6 end overlaps Sanger clone AL3548843; T7 end overlaps Sanger clone AL353619. We are only submitting the final 143.9 kilobases at the T7 end.

FEATURES

```

source          location/Qualifiers
1..143857
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/map="9"
/clone="RP11-1026H23"
/clone_id="RPC1-11 Human Male BAC"
19..278
/rpt_family="HAL1b"
complement(451..789)
/rpt_family="MLT1A1"
932..942
/note="<30 qual SNGL region"
959..965
/note="<30 qual SNGL region"
986..998
/note="<30 qual SNGL region"
1005..1009
/note="<30 qual SNGL region"
1059..1063
/note="<30 qual SNGL region"
1108..1114
/note="<30 qual SNGL region"
1133..1142
/note="<30 qual SNGL region"
1162..1167
/note="<30 qual SNGL region"
/note="<30 qual SNGL region"
complement(2267..2426)
/rpt_family="MER21B"
2484..2913
/rpt_family="MER31B"
complement(2950..3122)
/rpt_family="LTR3A"
3625..3663
/rpt_family="AT_rich"
4641..4683
/rpt_family="AT_rich"
complement(5259..5455)
/rpt_family="MIR"
5537..5580
/rpt_family="AT_rich"
complement(6557..6711)
/rpt_family="FRAM"
complement(7085..7256)
/rpt_family="MIR"
7822..7978
/rpt_family="MIR"
complement(8002..8427)
/rpt_family="LTR56"
11612..11786
/rpt_family="(GAAA)n"
11821..12214
/rpt_family="MLT1J2"
complement(12574..13349)
/rpt_family="L1PA16"
13356..14544
/rpt_family="L1P"
complement(14548..14730)
/rpt_family="L1MEC"
complement(14800..15265)
/rpt_family="L1MEC"
complement(15337..15444)

```

```

repeat_region /rpt_family="L1MEC"
15719..15752
/rpt_family="AT_rich"
repeat_region complement(15818..16111)
/rpt_family="AluSq"
repeat_region complement(16125..16404)
/rpt_family="AluSc"
/rpt_family="AluSc"
16972..16977
/note="<30 qual SNGL region"
16983..16991
/note="<30 qual SNGL region"
17075..17080
/note="<30 qual SNGL region"
/note="<30 qual SNGL region"
complement(17596..18051)
/rpt_family="L1M4"
complement(18086..18163)
/rpt_family="L1MA10"
18177..18527
/rpt_family="L1MDA"
18528..18671
/rpt_family="L1MC/D"
20646..20674
/rpt_family="AT_rich"
complement(21399..21794)
/rpt_family="MER11C"
complement(23176..23259)
/rpt_family="MIR"
complement(23495..23612)
/rpt_family="L1M4"
23613..23979
/rpt_family="HEIC"
complement(23980..24091)
/rpt_family="L1M4"
25843..26169
/rpt_family="L1"
26210..26378
/rpt_family="L1"
26648..27246
repeat_region

```

Query Match 66.1%; Score 21.8; DB 9; Length 143857;
 Best Local Similarity 92.0%; Pred. No. 31;
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 9 CATGCCGACAGAGTAACGATTAC 33
 Db 99216 CATGCTGACGACGATTACGATTAC 99192

Search completed: March 23, 2003, 16:00:35
 Job time : 264.806 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:37 ; Search time 38.371 Seconds
(without alignments)
1936.775 Million cell updates/sec

Title: US-09-873-546-2

Perfect score: 33

Sequence: 1 gggggatccatgcgcgcgaacagatcaacgattac 33

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_101002:.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	75.8	1249	24	ABQ72525
2	25	75.8	3061	24	ABQ72648
3	21.8	66.1	1087	22	AA160838
4	21.8	66.1	1108	22	AA159052
5	21.8	66.1	4167	24	AA037605
6	20.2	61.2	402	22	AA191981
7	20.2	61.2	459	21	AA022299
8	20.2	61.2	3249	21	AA078063
9	20.2	61.2	4011	23	AA594401

C	10	19.8	60.0	2187	22	AAE81507	White lupine secre
C	11	19.4	58.8	249487	24	ABN85733	Mouse genomic reg1
C	12	19.2	58.2	2081	21	AA078339	Human cancer assoc
C	13	19.2	58.2	5201	22	AAH90097	Human bone marrow
C	14	19.2	58.2	5294	22	AAH90071	Human bone marrow
C	15	19.2	58.2	5396	23	AA574120	DNA encoding novel
C	16	19.2	58.2	5404	24	AB162860	Breast cancer rela
C	17	19.2	58.2	5404	24	AB163081	Breast cancer rela
C	18	19.2	58.2	294458	24	ABN03041	Listeria monocytog
C	19	19.2	57.6	93	20	AAH85455	Human artery perfu
C	20	19.2	57.6	93	20	AAH85455	Human artery perfu
C	21	19.2	57.6	439	22	AA524757	Human ovarian PCR-
C	22	19.2	57.6	546	22	AAH83390	Human ovarian tumor
C	23	19.2	57.6	585	22	AAH11047	Human CDNA clone (
C	24	19.2	57.6	1424	22	AAH15075	Human CDNA sequenc
C	25	19.2	57.6	3266	21	AA090055	Murine Th2-specific
C	26	18.6	56.4	354	24	ABN67790	Streptococcus poly
C	27	18.6	56.4	1230	23	AB112537	Drosophila melanog
C	28	18.6	56.4	2252	23	AB123126	Drosophila melanog
C	29	18.6	56.4	3230	23	AB112536	Drosophila melanog
C	30	18.4	55.8	956	14	AA050152	Serratia marcescen
C	31	18.4	55.8	957	20	AA067274	Serratia spp. phos
C	32	18.4	55.8	957	21	AA256824	Serratia phospho11
C	33	18.4	55.8	1400	7	AAH60877	Sequence encoding
C	34	18.4	55.8	3001	21	AAH51782	Chromosome 13q31-q
C	35	18.4	55.8	3221	22	AAH14451	Human CDNA sequenc
C	36	18.2	55.2	1176	24	ABK63738	Rat sequence diffe
C	37	18.2	55.2	1716	24	ABQ90086	M. capsulatus gene
C	38	18.2	54.5	564	22	AAH81660	Human differential
C	39	18.2	54.5	2623	23	AB110171	Drosophila melanog
C	40	18.2	54.5	5326	23	AB118857	Drosophila melanog
C	41	18.2	54.5	3261	23	AB110170	Drosophila melanog
C	42	18.2	54.5	6288	22	AA544629	Fission yeast full
C	43	18.2	54.5	16424	22	AAK68448	Human immune/haema
C	44	18.2	54.5	16424	22	AAK68448	Human immune/haema
C	45	18.2	54.5	39651	23	AB118856	Drosophila melanog

ALIGNMENTS

RESULT 1	
ABQ72525	
ID	ABQ72525 standard; cDNA; 1249 BP.
XX	
AC	ABQ72525;
XX	
DT	03-SEP-2002 (first entry)
XX	
DE	Human MDDT encoding cDNA SEQ ID NO 77.
XX	
KW	Human: MDDT, disease detection and treatment molecule polynucleotide;
KW	proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
KW	autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
KW	rheumatoid arthritis; transgenic; gene therapy; antitumor; antitumor;
KW	hepatotoxic; antitumor; antitumor; antitumor; antitumor; antitumor;
KW	antitumor; antitumor; antitumor; antitumor; antitumor; antitumor;
XX	
OS	Homo sapiens.
XX	
PN	W0200240715-A2.
XX	
PD	23-MAY-2002.
XX	
PF	06-SEP-2001; 2001WO-US27628.
XX	
PR	06-SEP-2000; 2000US-230505P.
XX	
PR	06-SEP-2000; 2000US-230514P.
XX	
PR	06-SEP-2000; 2000US-230515P.
XX	
PR	06-SEP-2000; 2000US-230517P.
XX	
PR	06-SEP-2000; 2000US-230518P.
XX	
PR	06-SEP-2000; 2000US-230519P.

CC polynucleotide comprising. Oligonucleotides and antibodies are useful for
CC detecting MDR1 in a sample or for assessing toxicity of a test compound,
CC in a diagnostic test for a condition or a disease associated with the
CC expression of MDR1 in a biological sample, for detecting (1) in a sample,
CC and for purifying (1) from a sample. A composition comprising (1), an
CC agonist or antagonist is useful for treating a disease or condition
CC associated with decreased or increased expression of functional MDR1.
CC (1) or (11) are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of MDR1, where the disorders are
CC selected from a cell proliferative disorder such as arteriosclerosis,
CC cirrhosis, hepatitis, psoriasis, and cancer and an
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
CC rheumatoid arthritis. (11) are useful for creating knockin humanised
CC animals or transgenic animals to model human diseases, in somatic or
CC germ-line gene therapy, to generate a transcript image of a tissue or cell
CC type, for detecting differences in the chromosomal location due to
CC translocation or inversion among normal, carrier or affected individuals
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences.

CC Sequence 3061 BP; 606 A; 840 C; 927 G; 688 T; 0 other;

CC Query Match 75.8%; Score 25; DB 24; Length 3061;

CC Best Local Similarity 84.8%; Pred. No. 0.18;

CC Matches 28; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

CC 1 GGGGATCATCGCGAGACAGATACGATTAC 33

CC 42 GCGGGAGATGCCGAGACAGATACGATTAC 74

CC RESULT 3

CC ID AAI60838 standard; cDNA; 1087 BP.

CC AC AAI60838;

CC DT 22-OCT-2001 (first entry)

CC XX Human polynucleotide SEQ ID NO 4827.

KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.

XX Homo sapiens.

XX MO20015312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000MO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI, 2001-442253/47.

DR P-PSDB; AAM41682.

XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 4827; 10078bp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful

CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous

CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

CC utilisation of the activities such as: Immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

CC assays for receptor activity, arthritis and inflammation, leukaemias and

CC C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed

CC Specification.

CC Sequence 1087 BP; 270 A; 299 C; 321 G; 197 T; 0 other;

CC Query Match 66.1%; Score 21.8; DB 22; Length 1087;

CC Best Local Similarity 92.0%; Pred. No. 4.4;

CC Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

CC 9 CATGCCGAGACAGATACGATTAC 33

CC 150 CATGCCGAGACAGATACGATTAC 174

CC RESULT 4

CC ID AAI59052 standard; cDNA; 1108 BP.

CC AC AAI59052;

CC DT 22-OCT-2001 (first entry)

CC XX Human polynucleotide SEQ ID NO 1255.

KW Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.

XX Homo sapiens.

XX MO20015312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000MO-US34263.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX DR WPI, 2001-442253/47.

PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR P-PSDB; AAM3896.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 1255; 10078bp; English.
PS
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AA42213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 1108 BP; 274 A; 306 C; 328 G; 200 T; 0 other;
XX
Query Match 66.1%; Score 21.8; DB 22; Length 1108;
Best Local Similarity 92.0%; Pred. No. 4.4;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
-QY 9 CATGCCGACAGAGTACGATTAC 33
DB 171 CATGCCGACAGAGTACGATTAC 195
XX
-RESULT 5
AAD37605
XX AAD37605 standard; cDNA; 4167 BP.
XX
AC AAD37605;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human intracellular signalling (INTSIG-16) cDNA.
XX
XX Human; intracellular signalling; INTSIG-16; gene therapy; nausea; asthma;
XX Crohn's disease; keratosis; cell proliferative disorder; Grave's disease;
XX arteriosclerosis; hepatitis; leukaemia; autoimmune disorder; meningitis;
XX acquired immunodeficiency syndrome; AIDS; allergy; dermatitis; psoriasis;
XX Hashimoto's thyroiditis; neurological disorder; developmental disorder;
XX gastrointestinal disorder; reproductive disorder; neurotropic; antiemetic;
XX Alzheimer's disease; dementia; Creutzfeldt-Jacob disease; dermatological;
XX antiinflammatory; anaemia; nephrotropic; immunosuppressive; thyromimetic;
XX cancer; neuroprotective; cerebroprotective; antibacterial; virucide;
XX gynaecological; gene; ss.
XX
OS Homo sapiens.
XX
XX
FH Key Location/Qualifiers
FT CDS 161..760
FT /*tag= a
FT /product= "Human INSIG-16 protein"
FT sig_peptide 161..259
FT /*tag= b
FT mat_peptide 260..757
FT /*tag= c
FT /product= "Mature INTSIG-16 protein"
XX
XX WO200231152-A2.
XX
XX

PD 18-APR-2002.
XX
XX 12-OCT-2001; 2001WO-US32090.
PF
XX
XX 13-OCT-2000; 2000US-240871P.
PR 30-OCT-2000; 2000US-244723P.
PR 15-NOV-2000; 2000US-249402P.
PR 22-NOV-2000; 2000US-252622P.
PR 13-DEC-2000; 2000US-255622P.
XX
XX (INCYTE GENOMICS INC.
PA
XX Baughn MR, Ding L, Elliott VS, Gandhi AR, Gietzen KJ, Griffin JA;
XX Gururajan R, Hatalla AJA, Kearney L, Khan FA, Lal P, Lee EA;
XX Lu DAM, Lu Y, Nguyen DB, Arvizu C, Ramkumar J, Tang YT;
XX Thangavelu K, Thornton M, Walla NK, Warren BA, Xu Y, Yao MG;
XX Yue H;
XX WPI: 2002-426285/45.
DR P-PSDB; AAE23389.
XX
XX Novel intracellular signalling polypeptide useful for treating and
PT preventing cell proliferative, autoimmune, neurological,
PT gastrointestinal and reproductive and developmental disorders -
XX
XX Claim 91; Page 171-172; 177pp; English.
XX
XX The invention relates to an isolated intracellular signalling (INTSIG)
CC polypeptide or a biologically active or immunogenic fragment of INTSIG.
CC INTSIG is also useful for preparing a polyclonal or monoclonal antibody
CC by hybridoma technology. INTSIG gene is useful in gene therapy. INTSIG
CC is useful for treating or preventing disorders such as cell proliferative
CC disorder e.g. keratosis, arteriosclerosis, hepatitis, psoriasis,
CC leukaemia; autoimmune disorders e.g. AIDS, allergies, anaemia, asthma,
CC Crohn's disease, dermatitis, Graves' disease, Hashimoto's thyroiditis;
CC neurological disorders e.g. Alzheimer's disease, dementia, meningitis,
CC Creutzfeldt-Jacob disease; gastrointestinal disorders e.g. nausea,
CC reproductive disorders e.g. menstrual disorders, ectopic pregnancy and
CC cancer of the breast and developmental disorders. The present sequence
CC is human INTSIG-16 cDNA.
XX
XX
SQ Sequence 4167 BP; 1098 A; 911 C; 960 G; 1198 T; 0 other;
XX
Query Match 66.1%; Score 21.8; DB 24; Length 4167;
Best Local Similarity 92.0%; Pred. No. 5.3;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
-QY 9 CATGCCGACAGAGTACGATTAC 33
DB 160 CATGCCGACAGAGTACGATTAC 184
XX
-RESULT 6
AA191981C
XX AA191981 standard; cDNA; 402 BP.
XX
AC AA191981;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 12041.
DE
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200164835-A2.
XX
XX 07-SEP-2001.
XX
XX


```

PF 26-FEB-2001; 2001MO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YF, Liu C, Drmanac RT;
XX
DR WPI: 2001-514838/56.
DR P-PSDB: AAO12050.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosis and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 12041; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AA179941-AA193841) and
CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 402 BP; 88 A; 99 C; 127 G; 88 T; 0 other;
XX
SQ
XX
Query Match 61.2%; Score 20.2; DB 22; Length 402;
Best Local Similarity 75.8%; Pred. NO. 20;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0
XX
QY 1 GGGGGATCCATGCCGAGACAGACTAAGATTAC 33
II | | | | | | | | | | | | | | | | | |
Db 331 GGCAGCTCCATGCTGGAACAGACTAGGAAATTC 299
XX
RESULT 7
AAC02299/c
ID AAC02299 standard; cDNA: 459 BP.
XX
AAC02299;
XX
PT 06-OCT-2000 (first entry)
XX
XX Human secreted protein 5' EST, SEQ ID NO: 2297.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
XX gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
XX EP1033401-A2.
XX
XX 06-SEP-2000.
XX
XX 21-FEB-2000; 2000EP-0200610.
XX
XX 26-FEB-1999; 99US-0122487.
XX
XX (GEST ) GENSET.
XX
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX
XX WPI: 2000-500381/45.
XX
XX P-PSDB: AAG02293.

```

XX	New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT	obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX	diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX	Claim 1; SEQ ID 2297; 71pp + CD-ROM; English.
CC	The present sequence is one of a large number of 5' ESTs derived from
CC	mRNAs encoding secreted proteins. An ORF has been identified within the
CC	sequence. The 5' ESTs were prepared from total human RNAs or polyA ⁺ RNAs
CC	derived from 30 different tissues. EST sequences usually correspond
CC	mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC	often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC	well suited for isolating cDNA sequences derived from the 5' ends of
CC	mRNAs and even in those cases where longer cDNA sequences have been
CC	obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC	mRNAs with intact 5' ends and can therefore be used to obtain full length
CC	cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC	gene therapy and chromosome mapping procedures. They are used to obtain
CC	upstream regulatory sequences and to design expression and secretion
CC	vectors.
XX	
SEQ	Sequence 459 BP; 102 A; 105 C; 145 G; 106 T; 1 other:
Query Match	61.2%; Score 20.2; DB 21; Length 459;
Best Local Similarity	75.8%; Pred. No. 21;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;	
QY	1 GGGGATCCATGCGGAGACAGATGAACGATTAC 33
Db	325 GGCACCTCCATCTGGAACAGATAGGCAATTC 293
RESULT 8	
AAC78063	
ID	AAC78063 standard; cDNA; 3249 BP.
XX	
XX	AAC78063;
XX	
DT	08-FEB-2001 (first entry)
XX	
DE	Human cancer associated gene sequence SEQ ID NO:457.
XX	
KM	Human; cancer associated gene; cancer antigen; detection; cancer;
KM	diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KM	antidiabetic; antiasthmatic; antihemetic; antitachycardic; antiviral;
KM	antitubercular; antitumor; antiparasitic; antibacterial; cardiac;
KM	dermatologic; neuroprotective; thrombolytic; coagulant; nootropic;
KM	vasotropic; antipsoritic; antiangiogenic; gene therapy; inflammation;
KM	immune disorder; haematopoietic cell disorder; autoimmune disorder;
KM	allergic reaction; graft versus host disease; organ rejection;
KM	haemostatic; thrombolytic; cardiovascular disorder; infection;
KM	neurological disease; drug screening; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200055350-A1.
XX	
PD	21-SEP-2000.
XX	
PF	08-MAR-2000; 2000WO-US05882.
XX	
PR	12-MAR-1999; 99US-0124270.
XX	
PA	(HUMA-) HUMAN GENOME SCT INC.
XX	
PI	Rosen CA, Ruben SM;
XX	
DR	WPI: 2000-587533/55.
XX	
DR	P-PSDB; AAB43854.
XX	
PT	Novel isolated nucleic acids comprising sequences encoding peptides
PT	useful for treating or diagnosing e.g. cancer -

XX Claim 1; Page 993-994; 2352zp; English.

CC AAC77607 to AAC78448 encode the human cancer associated proteins given
XX In AAB3398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerary; immunomodulator;
CC antiinflammatory; antithyroid; antineumatic; antihairitic;
CC dermatological; neuroprotective; cardianc; thrombolytic; coagulant;
CC neotropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammatory, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.

XX Sequence 3249 BP; 900 A; 711 C; 748 G; 889 T; 1 other;

SQ

Query Match 61.2%; Score 20.2; DB 21; Length 3249;
Best Local Similarity 75.8%; Pred. No. 27;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0

OY 1 GGGGATCCTATGCCGAGACAGACTTACCATTC 33
|| ||||||| ||||||| || |
db 3001 GGCACTCCTCATCTGGAAACAGAGTAGGAATTC 3033

RESULT 9
AAS94401
ID AAS94401 standard; cDNA: 4011 BP.

XX AAS94401;
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #30205.
XX DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX P-PSTDB: ABG30214.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XS Claim 1; SEQ ID NO 30205; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (III) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 4011 BP; 1075 A; 910 C; 953 G; 1073 T; 0 other;

Query Match 61.2%; Score 20.2; DB 23; Length 4011;
Best Local Similarity 75.8%; Pred. No. 28;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DY 1 GGGGATCCATGCCCGACAGACTAACCATTAC 33
||| | ||||||| ||||||||||| ||| |
Db 3099 GGCACTCCTCATGTGGAACAGAGTAGGAAATTTC 3131

RESULT 10
AAF81507/C
ID AAF81507 standard; DNA; 2187 BP.

XX AAF81507;
AC
XX
DT 05-JUN-2001 (first entry)
XX
DE White lupine secretory acid phosphatase coding sequence.
KW White lupine: secretory acid phosphatase; root; phosphor; ds.
XX
OS Lupinus albus.
XX
FH Location/Qualifiers
FT CDS 1..1917
FT /*tag= a
FT /product= "secretory acid phosphatase"

PN JP2001029085-A.
XX
PD 06-FEB-2001.
XX
PF 07-JAN-2000; 2000JP-0001442.
PR 18-MAY-1999; 99JP-0137273.
XX
PA (MITA) MITSUI CHEM INC.
XX
DR WPI: 2001-221499/23.
P-PSDB: AAB74496.

PT Secretory acid phosphatase -
XX
XX
PS Claim 1; Page 9-11; 17pp; Japanese.

CC The present invention provides the protein and coding sequences of a
CC secretory acid phosphatase from the white lupine which combines to the
CC surface of plant roots. The sequences are useful for producing plants

CC AAC7760 to AAC78448 encode the human cancer associated proteins given
CC in ABA43398 to ABA44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cyrostatic; proliferative; vlnerary; immunomodulator;
CC antidiabetic; antilasthmatic; antihemmatic; antiautritic;
CC antinflmmatory; antihypoid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC neotropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and ABA44240 represent sequences used in the exemplification of
CC the present invention.

Query Match	58.28;	Score 19.2;	DB 22;	Length 5234;
Best Local Similarity	75.08;	Pred. No. 81;		
Matches 24;	Conservative	0;	Mismatches 8;	Indels 0;

QY	2	GGGGATCCATGCCGGAACAGAGTAACGATTAC	33
Db	4563	GGAGATCCACGCGAGTAAGTACTGTGAC	45322

RESULT 15
AAS74120/c

AC AAS74120;

DT 13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #9924.

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss

OS Homo sapiens.

PN W0200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Llu C, Tang YT,

DR WPI; 2001-639362/73.

DR P-PSDB; ABG09933.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutation
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 1; SEQ ID No 9924; 103pp; English

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving CC (II). (II) is useful for generating antibodies against it, detecting or quantifying a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. A564197-A594564 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 5396 BP; 1623 A; 1085 C; 1157 G; 1531 T; 0 other;

Query Match	58.28;	Score 19.2;	DB 23;	Length 5396;
-------------	--------	-------------	--------	--------------

Best Local Similarity 75.0%; Pred. No. 81;
Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY	1	G G G G G A T C C A T G C C C G A A C A G A G T A A C A T T A	32
Db	5214	G G G G C A T C C A G G T G G G A A C A T C G T A A G G A C T A	5183

```
Search completed: March 23, 2003, 15:19:51
Job time : 113.871 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:56 ; Search time 2058.86 Seconds

(without alignments)
8438.824 Million cell updates/sec

Title: US-09-873-546-4

Perfect score: 597

Sequence: 1 atgcgcgaacagagtaacga.....gcaaatgcaccctcatgtga 597

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: GenDbml:*
2: gb_da:*
3: gb_hg:*
4: gb_in:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	595.4	99.7	597	9	AY056037
2	595.4	99.7	597	9	AY059641
3	595.4	99.7	774	9	AB076888
4	595.4	99.7	2827	9	AK096600
5	595.4	99.7	3391	9	BC030660
6	595.4	99.7	177540	9	AC006538
7	583.4	97.7	1249	6	AX430295
8	582.8	97.6	3061	6	AX430418
9	501	83.9	597	10	AY059640
10	501	83.9	126901	2	AC091518
11	499.4	83.9	201320	2	AC073816
12	499.4	83.7	153386	2	AC103000
13	385.6	64.6	4412	9	AB062937
14	380.8	63.8	1875	9	BC008065
15	380.8	63.8	2505	9	AB076889
16	380.8	63.8	4167	6	AX477382
17	380.8	63.8	143299	9	AL353619
18	380.8	63.8	143857	9	AC021583
19	380.8	63.8	199612	9	AC021582
20	223.6	37.5	1971	9	AK096393
21	222	37.2	1286	6	AX077661
22	222	37.2	1495	9	HS096750
23	222	37.2	1511	9	BC005362
24	222	37.2	1561	9	AK021882
25	222	37.2	7242	9	AF202543
26	222	37.2	142396	9	AL157407
27	209.6	35.1	68052	2	AC100691
28	204	34.2	68052	2	AC100691
29	175.4	29.4	552	9	AF493915
30	175.4	29.4	552	9	HSRAP2B
31	175.4	29.4	1913	9	BC012362
32	175.4	29.4	140794	2	AC069415
33	175.4	29.4	165733	2	AC069072
34	175.4	29.4	198170	9	AC117394
35	175	29.3	2145	9	HSMB03027
36	174.2	29.2	240020	2	AC121587
37	174.2	29.2	289723	2	AC122935
38	167.4	28.0	552	10	AF386786
39	166.4	27.9	552	9	AF493914
40	166.4	27.9	555	9	HSRAP2
41	166.4	27.9	558	6	A08695
42	160.6	26.9	33443	2	AC014332
43	160.6	26.9	160817	3	AC008316
44	160.6	26.9	177028	3	AC008315
45	160.6	26.9	219579	3	AE003684

ALIGNMENTS

RESULT 1
AY056037 597 bp mRNA linear PRI 24-JUL-2002
LOCUS AY056037
DEFINITION Homo sapiens R1g protein mRNA, complete cds.
ACCESSION AY056037
VERSION AY056037.1 GI:16555333
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 597)
AUTHORS Ellis,C.A., Vos,M.D., Howell,H., Vallecora,T., Fults,D.W. and Clark,G.J.
TITLE R1g is a novel Ras-related protein and potential neural tumor

Pred. No. is the number of results predicted by chance to have a

JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

suppressor
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 9876-9881 (2002)
12107278
2 (bases 1 to 597)
Castro, A.F. and Quilliam, L.A.
Identification of R19, a novel Ras family member sharing homology
with NOY2 and Rap
Unpublished
3 (bases 1 to 597)
Castro, A.F. and Quilliam, L.A.
Direct Submission
Submitted (13-SEP-2001) Biochemistry and Molecular Biology, Indiana
University School of Medicine, 635 Barnhill Drive, MS-4053,
Indianapolis, IN 46202, USA
Location/Qualifiers
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13.3"
/tissue_type="brain; anaplastic oligodendroglioma"
/note="derived from dbEST AI497811"
1..597
/note="Ras family GTPase"
/codon_start=1
/product="R19 protein"
/protein_id="AA123715.1"
/db_xref="GI:1655334"
/translation="MPEOSNDYRVVFGAGVGSLSVLRFKYGFDPDYPTIEDTY
RQVTSCKDSVCTLOITDPTMGSHQFAPAMORLSKSHAPILVSVYSKSLDELGIYK
LIVQIKGSVEDIPVAVGNKCDREVDREDAQVADQWKAEMTSKKNYKEL
FOELLTLEFRMSLNDGKRSKOKRDRYKGTLM"

CDS
BASE COUNT 134 a 178 c 195 g 90 t
ORIGIN

Query Match 99.7%; Score 595.4; DB 9; Length 597;
Best Local Similarity 99.8%; Pred. No. 5.9e-82;
Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCCGGAACAGATTAACGATTACCGCGTGTGTTGCGGGCGGCGCGTGGGCAAG 60
DB 1 ATGCCGGAACAGATTAACGATTACCGCGTGTGTTGCGGGCGGCGCGTGGGCAAG 60
QY 61 AGCTGCTGTGCTGCTGCTTCTGGAAGGACGCTTCCGGACACTTACATCCCAACATC 120
DB 61 AGCTGCTGTGCTGCTGCTTCTGGAAGGACGCTTCCGGACACTTACATCCCAACATC 120
QY 121 GAGGACACTTACCGGAGGATGATCGACGAAGAGGCTGTGACGCTGCAGATCACA 180
DB 121 GAGGACACTTACCGGAGGATGATCGACGAAGAGGCTGTGACGCTGCAGATCACA 180
QY 181 GACACACCGGAGCAGCAGCAGTTCGCCGCAATGACAGCGCTGTCTTCCAAAGGGCCAC 240
DB 181 GACACACCGGAGCAGCAGCAGTTCGCCGCAATGACAGCGCTGTCTTCCAAAGGGCCAC 240
QY 241 GCTTCATCTGTGTTCTTCCGTCACACAGCAAGCACTGAGAGAGCTGGGGCCATTC 300
DB 241 GCTTCATCTGTGTTCTTCCGTCACACAGCAAGCACTGAGAGAGCTGGGGCCATTC 300
QY 301 TCAACACTATGTCAGATCAAGGAGGAGGAGATCCCGCTGATGCTCGTGGGC 360
DB 301 TCAACACTATGTCAGATCAAGGAGGAGGAGATCCCGCTGATGCTCGTGGGC 360
QY 361 AACCAAGTGGATGAGACGAGCGGAGTGTGACACGCGGAGGCGGAGCGGTGGCCAA 420
DB 361 AACCAAGTGGATGAGACGAGCGGAGTGTGACACGCGGAGGCGGAGCGGTGGCCAA 420
QY 421 GAGTGAAGTGGCTTTTCTTGAAGACCTGCGCAAGATTAACATCAAGCTCAAGAGCTC 480
DB 421 GAGTGAAGTGGCTTTTCTTGAAGACCTGCGCAAGATTAACATCAAGCTCAAGAGCTC 480
QY 481 TTCCAGGAGCTGTGACGCTGAGAGCGCGGAGATGAGCTCCGCTGATGCTCGTGGGC 540
DB 481 TTCCAGGAGCTGTGACGCTGAGAGCGCGGAGATGAGCTCCGCTGATGCTCGTGGGC 540

DB 481 TTCCAGGAGCTGTGACGCTGAGAGCGCGGAGATGAGCTCCGCTGATGCTCGTGGGC 540
QY 541 CGCTCCGGAGAGAGAGAGAGAGACCGCGTCAAGGCAATGACCCCTCATGTGA 597
DB 541 CGCTCCGGAGAGAGAGAGAGAGACCGCGTCAAGGCAATGACCCCTCATGTGA 597

RESULT 2
AY059641
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

AY059641
597 bp mRNA linear PRI 29-OCT-2001
Homo sapiens small GTP-binding tumor suppressor 1 mRNA, complete
cds.
AY059641
AY059641.1 GI:16508175
Homo sapiens.
Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 597)
Gong, L. and Wu, K.
Molecular cloning of GBRSL, a novel gene encoding a small
GTP-binding tumor suppressor
2 (bases 1 to 597)
Unpublished
Gong, L.
Direct Submission
Submitted (12-OCT-2001) Dept. of Cardiology, U.T. MD. Anderson
Cancer Center, 1515 Holcombe Blvd. 449, Houston, TX 77030, USA
Location/Qualifiers
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
1..597
/note="GBRSL"
/codon_start=1
/product="small GTP-binding tumor suppressor 1"
/protein_id="AA117968.1"
/db_xref="GI:16508176"
/translation="MPEOSNDYRVVFGAGVGSLSVLRFKYGFDPDYPTIEDTY
RQVTSCKDSVCTLOITDPTMGSHQFAPAMORLSKSHAPILVSVYSKSLDELGIYK
LIVQIKGSVEDIPVAVGNKCDREVDREDAQVADQWKAEMTSKKNYKEL
FOELLTLEFRMSLNDGKRSKOKRDRYKGTLM"

CDS
BASE COUNT 134 a 178 c 195 g 90 t
ORIGIN

Query Match 99.7%; Score 595.4; DB 9; Length 597;
Best Local Similarity 99.8%; Pred. No. 5.9e-82;
Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGCCGGAACAGATTAACGATTACCGCGTGTGTTGCGGGCGGCGCGTGGGCAAG 60
DB 1 ATGCCGGAACAGATTAACGATTACCGCGTGTGTTGCGGGCGGCGCGTGGGCAAG 60
QY 61 AGCTGCTGTGCTGCTGCTTCTGGAAGGACGCTTCCGGACACTTACATCCCAACATC 120
DB 61 AGCTGCTGTGCTGCTGCTTCTGGAAGGACGCTTCCGGACACTTACATCCCAACATC 120
QY 121 GAGGACACTTACCGGAGGATGATCGACGAAGAGGCTGTGACGCTGCAGATCACA 180
DB 121 GAGGACACTTACCGGAGGATGATCGACGAAGAGGCTGTGACGCTGCAGATCACA 180
QY 181 GACACACCGGAGCAGCAGCAGTTCGCCGCAATGACAGCGCTGTCTTCCAAAGGGCCAC 240
DB 181 GACACACCGGAGCAGCAGCAGTTCGCCGCAATGACAGCGCTGTCTTCCAAAGGGCCAC 240
QY 241 GCTTCATCTGTGTTCTTCCGTCACACAGCAAGCACTGAGAGAGCTGGGGCCATTC 300
DB 241 GCTTCATCTGTGTTCTTCCGTCACACAGCAAGCACTGAGAGAGCTGGGGCCATTC 300
QY 301 TCAACACTATGTCAGATCAAGGAGGAGGAGATCCCGCTGATGCTCGTGGGC 360
DB 301 TCAACACTATGTCAGATCAAGGAGGAGGAGATCCCGCTGATGCTCGTGGGC 360


```

Db 301 TACAAGTCATCGTGCAGATCAAGGCGCGGTGAGACATCCCGTGATGCTGTGGGC 360
QY 361 AACAAGTCGATGAGACGACGCGGAGGTGAGACGCGGAGGCGAGGGGCGGCCAA 420
Db 361 AACAAGTCGATGAGACGACGCGGAGGTGAGACGCGGAGGCGAGGGGCGGCCAA 420
QY 421 GAGTGAAGTGCCTTTCATGAGACCTCGGCCAAGTGAACATCAACATCAAGGAGCTC 480
Db 421 GAGTGAAGTGCCTTTCATGAGACCTCGGCCAAGTGAACATCAACATCAAGGAGCTC 480
QY 481 TTCAGAGAGCTGCTGACGCTGAGACGCGCGGAACTGAGCTCAACATCGAGCGCAAG 540
Db 481 TTCAGAGAGCTGCTGACGCTGAGACGCGCGGAACTGAGCTCAACATCGAGCGCAAG 540
QY 541 CGCTCCGGAAGCAGAGAGACGACGCGGTCAAGGGCAATGACCATCTCATGTGA 597
Db 541 CGCTCCGGAAGCAGAGAGACGACGCGGTCAAGGGCAATGACCATCTCATGTGA 597

RESULT 3
LOCUS AB076888 774 bp mRNA linear PRI 25-JUN-2002
DEFINITION Homo sapiens mRNA for Di-Ras1, complete cds.
ACCESSION AB076888
VERSION AB076888.1 GI:21624247
KEYWORDS Homo sapiens adult brain cDNA to mRNA, clone_lib:lambda ZAPII human brain cDNA.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1
AUTHORS Kontani,K., Ogawa,T., Okai,T., Tada,M. and Katada,T.
TITLE Di-Ras: A Distinct Subgroup of Ras-family GTPases with Unique Biochemical Properties
JOURNAL Unpublished
2 (bases 1 to 774)
AUTHORS Kontani,K., Ogawa,T., Okai,T., Tada,M. and Katada,T.
TITLE Direct Submission
JOURNAL Submitted (24-DEC-2001) Toshiaki Katada, University of Tokyo, Department of Physiological Chemistry, Graduate School of Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan (E-mail:katadaemol.f.u-tokyo.ac.jp, Tel:81-3-5841-4750, Fax:81-3-5841-4751)
FEATURES
source location/Qualifiers
1..774
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13.3"
/tissue_type="brain"
/clone_lib="lambda ZAPII human brain cDNA"
/dev_stage="adult"
1..774
/gene="Di-Ras1"
/locus_tag="AB076888"
/feature_key="CDS"
/translation="MPEQSDRYVYVFGAGVSKSLVLFYKGTFRDRTYPIPTEDTY
KQVISCDSKSVCTIQTIDTIGSHOFPAMQRSLISKAFILVTSKQSLFEEGLPYK
LIVQIGSVEDIPVLMVGNKCDTQREVTREQAQAAQEWKCAFMTSAAKMNVAEKL
FQELTLERNNRSLNIDKRSKQKRTREVRVKKCTLM"
BASE COUNT 154 a 268 c 239 g 113 t
ORIGIN
Query Match 99.78; Score 595.4; DB 9; Length 774;
Best Local Similarity 99.88; Pred. No. 5.6e-82;
Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 ATGCGGAGACAGATACGATTACCGCGTGTGTGTTGGGGGCGGCGGTGGCAAG 60
Db 100 ATGCGGAGACAGATACGATTACCGCGTGTGTGTTGGGGGCGGCGGTGGCAAG 159
QY 61 AGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
Db 160 AGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 219
QY 121 GAGGACACCTTACCGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180
Db 220 GAGGACACCTTACCGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 279
QY 181 GACACACCGGAGGACCGACAGTTCGCGCATGAGCGCTGTTCATCTCCAAAGGCGAC 240
Db 280 GACACACCGGAGGACCGACAGTTCGCGCATGAGCGCTGTTCATCTCCAAAGGCGAC 339
QY 241 GCCTTCATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
Db 340 GCCTTCATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 399
QY 301 TACAAGTCATCTGTCAGATCAAGGCGGAGTGTGAGGACATCCCGTATGCTGTGGGC 360
Db 400 TACAAGTCATCTGTCAGATCAAGGCGGAGTGTGAGGACATCCCGTATGCTGTGGGC 459
QY 361 AACAAGTCGATGAGACGACGCGGAGGTGAGACGCGGAGGCGAGGCGGTGGCCAA 420
Db 460 AACAAGTCGATGAGACGACGCGGAGGTGAGACGCGGAGGCGAGGCGGTGGCCAA 519
QY 421 GAGTGAAGTGCCTTTCATGAGACCTCGGCCAAGTGAACATCAACATCAAGGAGCTC 480
Db 520 GAGTGAAGTGCCTTTCATGAGACCTCGGCCAAGTGAACATCAACATCAAGGAGCTC 579
QY 481 TTCAGAGAGCTGCTGACGCTGAGACGCGCGGAACTGAGCTCAACATCGAGCGCAAG 540
Db 580 TTCAGAGAGCTGCTGACGCTGAGACGCGCGGAACTGAGCTCAACATCGAGCGCAAG 639
QY 541 CGCTCCGGAAGCAGAGAGACGACGCGGTCAAGGGCAATGACCATCTCATGTGA 597
Db 640 CGCTCCGGAAGCAGAGAGACGACGCGGTCAAGGGCAATGACCATCTCATGTGA 696

RESULT 4
LOCUS AK096600 2827 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ39281 fis, clone OCBFP2011067, highly similar to Homo sapiens cingulin mRNA.
ACCESSION AK096600
VERSION AK096600.1 GI:21756131
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens fetal brain cDNA to mRNA, clone_lib:OCBBF2 clone:OCBBF2011067.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1
AUTHORS Nishi,T., Oka,T., Nakagawa,S., Senoh,A., Mizunuchi,H., Inagaki,H., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai,Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahara,K., Masuno,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
2 (bases 1 to 2827)
AUTHORS Isogai,T. and Yamamoto,J.
JOURNAL Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing:

```



```

repeat_region      /rpl_family="AluSx"
                   complement(868. .931)
                   /rpl_family="L1MB7"
                   1175. .1365
                   /note="Predicted exon, program: grrail2exons_human_1.3
                   frame: 2, quality: good, score: 57.000"
repeat_region      /rpl_family="AluSg"
                   complement(2133. .2231)
                   /rpl_family="MER81"
                   2739. .2935
repeat_region      /rpl_family="AluJo"
                   2946. .3247
                   /rpl_family="AluSx"
                   3256. .3384
repeat_region      /rpl_family="L1MC3"
                   3389. .3604
repeat_region      /rpl_family="AluJo"
                   3631. .3716
                   /rpl_family="L1MD3"
                   complement(4087. .4242)
repeat_region      /rpl_family="AluJo"
                   complement(4294. .4436)
repeat_region      /rpl_family="AluJo"
                   complement(4598. .4887)
                   /rpl_family="AluSg"
                   complement(4906. .5373)
repeat_region      /rpl_family="L1MB7"
                   complement(5382. .5669)
                   /rpl_family="AluSx"
                   complement(6425. .6609)
repeat_region      /rpl_family="GGCA)n"
                   complement(6610. .6765)
repeat_region      /rpl_family="(GAA)n"
                   complement(6766. .7065)
repeat_region      /rpl_family="AluSx"
                   7093. .7392
repeat_region      /rpl_family="AluSx"
                   7423. .7511
repeat_region      /rpl_family="L1MA7"
                   7539. .7839
repeat_region      /rpl_family="AluSx"
                   7852. .7938
                   /rpl_family="L1MA7"
repeat_region      8001. .8312
                   /rpl_family="AluJo"
repeat_region      complement(8711. .8897)
                   /rpl_family="MER20"
repeat_region      complement(9532. .9722)
                   /rpl_family="MER20"
repeat_region      complement(9970. .10143)
                   /rpl_family="MER20"
repeat_region      10144. .10405
                   /rpl_family="AluJo"
repeat_region      10421. .10630
                   /rpl_family="L1M3"
repeat_region      10631. .10756
                   /rpl_family="AluSg"
repeat_region      10757. .11058
                   /rpl_family="AluV"
repeat_region      11068. .11243
                   /rpl_family="AluSg"
repeat_region      11247. .11520
                   /rpl_family="L1M3"
repeat_region      11522. .11816
                   /rpl_family="AluSg"
repeat_region      complement(11827. .12033)
                   /rpl_family="MER58A"
repeat_region      12255. .12555
                   /rpl_family="AluJo"
repeat_region      complement(12562. .12617)
                   /rpl_family="(TA)n"
repeat_region      12769. .12894

```

Query Match	Best Local Similarity	Matches	Score	Pred. No.	DB	Length
99.7%	99.8%	596	595.4	2.2e-82	177540	0
Conservative	0	Mismatches	1	Indels	0	Gaps
0	0	0	0	0	0	0

repeat_region	repeat_family="Alus9"	complement(24739, .24926)	repeat_region	repeat_family="MER58A"	complement(24739, .24926)
repeat_region	repeat_family="Alus9"	complement(24739, .24926)	repeat_region	repeat_family="MER58A"	complement(24739, .24926)
repeat_region	repeat_family="Alus9"	complement(24739, .24926)	repeat_region	repeat_family="MER58A"	complement(24739, .24926)

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	JOURNAL	FEATURES
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX430295	GI:21655659	human.	1	Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.	Patent: WO 0240715-A 77 23-MAY-2002.	Location/Qualifiers
AX430295	Sequence 77 from Patent WO0240715.	AX430295	AX4						


```

/gene="gbts1"
/locus="gbts1"
/codon_start=1
/product="small GTP-binding tumor suppressor 1"
/db_xref="GI:16508174"
/translation="MPEOSNDYRVVVFAGVGVKSSLYLRFKGFEDRDTPTIEDTY
RQVIGSKDSVCTPLQITDTGSHQFPAMOLSTSKHAFILFVSYSKSLDLSPIYK
LIVQIGSVEDIPIMLVGNKCDERQREVTREARQAVADPMKAFETSKMNYNVEL
FOELTLETRRSVSVGDKRSSKORADRIKGCALM"
BASE COUNT      138 a      166 c      194 g      99 t
ORIGIN
Query Match      83.9%; Score 501; DB 10; Length 597;
Best Local Similarity 89.9%; Pred. No. 1.7e-67;
Matches 537; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 ATGCCGAACAGATACATTAACCGCGTGTGTGCGGGCGGCGCGTGGCCAAAG 60
DB 1 ATGCCGAACAGAGCAATACCTGCTGTGTGTGCGGGCGGCGCGTGGCCAAAG 60
QY 61 AGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
DB 61 AGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
QY 121 GAGGACACCTACCGGAGGTGATCAGCTGCGACAAAGAGCGGTGACGTGACATCACA 180
DB 121 GAGGACACCTACCGGAGGTGATCAGCTGCGACAAAGAGCGGTGACGTGACATCACA 180
QY 181 GACACACCGGAGGACGACAGTTCCTCCGCGCATGACGCGCTGTCCATCTCAAGGGCCAC 240
DB 181 GACACACCGGAGGACGACAGTTCCTCCGCGCATGACGCGCTGTCCATCTCAAGGGCCAC 240
QY 241 GCCTTATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
DB 241 GCCTTATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
QY 301 TACAAAGCTCATGTCAGATCAAGGCGACAGCTGTGAGAGACATCCCGTGTCTGTGGGC 360
DB 301 TACAAAGCTCATGTCAGATCAAGGCGACAGCTGTGAGAGACATCCCGTGTCTGTGGGC 360
QY 361 AACAGTGTGATGATGACGAGCGGAGGAGTGTGACACGCGGAGCGCGAGCGGTGGCCCA 420
DB 361 AACAGTGTGATGATGACGAGCGGAGGAGTGTGACACGCGGAGCGCGAGCGGTGGCCCA 420
QY 421 GAGTGAAGTGGCTTTTATGAGACCTGCGGCAAGATGAACTAAGCTCAAGAGAGCTC 480
DB 421 GAGTGAAGTGGCTTTTATGAGACCTGCGGCAAGATGAACTAAGCTCAAGAGAGCTC 480
QY 481 TTCCAGAGAGCTGTGACGCTGTGAGAGCGCGCGGACATGAGCTTCACATCGAGCGCAA 540
DB 481 TTCCAGAGAGCTGTGACGCTGTGAGAGCGCGCGGACATGAGCTTCACATCGAGCGCAA 540
QY 541 CGCTCCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 597
DB 541 CGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 597

RESULT 10
LOCUS      AC091518      126901 bp      DNA      linear      HTG 14-AUG-2002
DEFINITION Mus musculus chromosome 10 clone mgs1-257j2 strain 129/svj, WORKING
ACCESSION  AC091518
VERSION     AC091518.11 GI:22218471
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 126901)
AUTHORS    Kim,J., Shauli,S., Yao,Z., Andrews,G.K. and Roe,B.A.
TITLE      Mus musculus BAC Clone mgs1-257j2

```

```

JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 126901)
AUTHORS      Kim,J., Shauli,S., Yao,Z., Andrews,G.K. and Roe,B.A.
TITLE        Direct Submission
JOURNAL      Submitted (29-APR-2001) Department of Chemistry And Biochemistry,
              The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
REFERENCE    3 (bases 1 to 126901)
AUTHORS      Kim,J., Shauli,S., Yao,Z., Andrews,G.K. and Roe,B.A.
TITLE        Direct Submission
JOURNAL      Submitted (14-AUG-2002) Department of Chemistry And Biochemistry,
              The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
              OK 73019, USA
COMMENT      On Aug 14, 2002 this sequence version replaced gi:21955024.
              ----- Genome Center
              Center: Department of Chemistry And Biochemistry
              The University of Oklahoma
              Center code:UOKNOR
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 5 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1
              * 7755: contig of 7755 bp in length
              * 7756: gap of unknown length
              * 7856: contig of 21593 bp in length
              * 29448: gap of unknown length
              * 29449: contig of 29277 bp in length
              * 58825: gap of unknown length
              * 58826: gap of unknown length
              * 58925: contig of 30693 bp in length
              * 58926: gap of unknown length
              * 89619: contig of 37183 bp in length.
              * 89718: gap of unknown length
              * 89719: contig of 37183 bp in length.
              *
              Location/Qualifiers
                source
                  1. 126901
                     /organism="Mus musculus"
                     /strain="129/Svj"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /clone="mgs1-257j2"
                     /clone_11b="Genome Systems mouse ES cell BAC libraries I
                     and II"
BASE COUNT   30097 a 33386 c 32214 g 30700 t      504 others
ORIGIN
Query Match      83.9%; Score 501; DB 2; Length 126901;
Best Local Similarity 89.9%; Pred. No. 6.5e-68;
Matches 537; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 ATGCCGAACAGATACATTAACCGCGTGTGTGCGGGCGGCGCGTGGGCAAG 60
DB 26982 ATGCCGAACAGAGCAATACCTGCTGTGTGTGCGGGCGGCGCGTGGGCAAG 27041
QY 61 AGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
DB 27042 AGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
QY 121 GAGGACACCTACCGGAGGTGATCAGCTGCGACAAAGAGCGGTGACGTGACATCACA 180
DB 27102 GAGGACACCTACCGGAGGTGATCAGCTGCGACAAAGAGCGGTGACGTGACATCACA 180
QY 181 GACACACCGGAGGACGACAGTTCCTCCGCGCATGACGCGCTGTCCATCTCAAGGGCCAC 240
DB 27162 GACACACCGGAGGACGACAGTTCCTCCGCGCATGACGCGCTGTCCATCTCAAGGGCCAC 27221
QY 241 GCCTTATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
DB 27222 GCCTTATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
QY 301 TACAAAGCTCATGTCAGATCAAGGCGACAGCTGTGAGAGACATCCCGTGTGTGGGC 360

```


AC103000/c 153386 bp DNA linear HTG 12-JUL-2002
LOCUS Rattus norvegicus clone CH230-192M3, *** SEQUENCING IN PROGRESS
DEFINITION *** 63 unordered pieces.
ACCESSION AC103000 GI:21729632
VERSION HTG: HTGS_PBASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 153386)
Muzny,D.M., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbera,J., Benton,J., Binnage,K., Blakenburg,K., Bonnin,D.,
Bouck,U., Bowle,S., Brlewa,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
DeValla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Deaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homi,F., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lozdo,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogun,M., Okunolu,G.,
Oetungye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Syatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G., and Gibbs,R.
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GICZ
Center clone name: CH230-192M3
----- Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap version 0.990329
Consensus quality: 83670 bases at least Q40
Consensus quality: 86659 bases at least Q30
Consensus quality: 88900 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1030 1029: contig of 1029 bp in length
1130 1129: gap of unknown length
2363 2363: contig of 1234 bp in length
2463 2463: gap of unknown length
3463 3463: contig of 1000 bp in length
3563 3563: gap of unknown length
5122 5122: contig of 1559 bp in length
5223 5223: gap of unknown length
5502 5502: contig of 1280 bp in length
6503 6502: gap of unknown length
6603 7941: contig of 1339 bp in length
7942 8041: gap of unknown length
8042 9124: contig of 1083 bp in length
9125 9224: gap of unknown length
9225 10411: contig of 1187 bp in length
10412 10511: gap of unknown length
10512 12121: contig of 1610 bp in length
12122 12221: gap of unknown length
12222 13926: contig of 1705 bp in length
13927 14026: gap of unknown length
14027 15525: contig of 1499 bp in length
15526 15625: gap of unknown length
15626 17072: contig of 1447 bp in length
17073 17172: gap of unknown length
17173 18485: contig of 1313 bp in length
18486 18585: gap of unknown length
18586 19821: contig of 1236 bp in length
19822 21254: gap of unknown length
21255 21354: contig of 1333 bp in length
21355 22467: contig of 1113 bp in length
22468 22567: gap of unknown length
22568 23903: contig of 1336 bp in length
23904 24003: gap of unknown length
24004 25789: contig of 1786 bp in length
25790 25889: gap of unknown length
25890 27325: contig of 1436 bp in length
27326 27425: gap of unknown length
27426 29320: contig of 1895 bp in length
29321 29420: gap of unknown length
29421 30862: gap of 1442 bp in length
30863 30962: gap of unknown length
30963 32929: contig of 1967 bp in length
32930 33029: gap of unknown length
33030 34612: contig of 1583 bp in length
34613 34712: gap of unknown length
34713 35663: contig of 1851 bp in length
35664 36663: gap of unknown length
36664 37847: contig of 1184 bp in length
37848 39747: gap of unknown length
39749 39847: contig of 1800 bp in length
39847: gap of unknown length


```

* 39848 40864: contig of 1017 bp in length
* 40865 40964: gap of unknown length
* 42896 42966: contig of 1932 bp in length
* 42897 42996: gap of unknown length
* 42997 44210: contig of 1214 bp in length
* 44211 44310: gap of unknown length
* 44311 45381: contig of 1071 bp in length
* 45382 45481: gap of unknown length
* 45482 47015: contig of 1534 bp in length
* 47016 47115: gap of unknown length
* 47116 48966: contig of 1881 bp in length
* 48967 49066: gap of unknown length
* 49067 50509: contig of 1413 bp in length
* 50510 50609: gap of unknown length
* 50610 51947: contig of 1338 bp in length
* 51948 52047: gap of unknown length
* 52048 54858: contig of 2811 bp in length
* 54859 54958: gap of unknown length
* 54959 56291: contig of 1333 bp in length
* 56292 56391: gap of unknown length
* 56392 57989: contig of 1598 bp in length
* 57990 58089: gap of unknown length
* 58090 59713: contig of 1624 bp in length
* 59714 59813: gap of unknown length
* 59814 61504: contig of 1691 bp in length
* 61505 62984: contig of 1380 bp in length
* 62985 63084: gap of unknown length
* 63085 64788: contig of 1704 bp in length
* 64789 64888: gap of unknown length
* 64889 66582: contig of 1694 bp in length
* 66583 66842: gap of unknown length
* 66843 68842: contig of 2160 bp in length
* 68843 70931: gap of unknown length
* 70932 71031: contig of 1989 bp in length
* 71032 73333: contig of 2302 bp in length
* 73334 73433: gap of unknown length
* 73434 75603: contig of 2170 bp in length
* 75604 75703: gap of unknown length
* 75704 78432: contig of 2729 bp in length
* 78433 78532: gap of unknown length
* 78533 80528: contig of 1996 bp in length
* 80529 80628: gap of unknown length
* 80629 83682: contig of 3054 bp in length
* 83683 83783: gap of unknown length
* 83783 87933: contig of 4151 bp in length
* 87934 88033: gap of unknown length
* 88034 91869: contig of 3836 bp in length
* 91870 94876: contig of 2907 bp in length
* 94877 94976: gap of unknown length
* 94977 100254: contig of 5278 bp in length
* 100255 100354: gap of unknown length
* 100355 105563: contig of 5209 bp in length

```

Query Match Best Local Similarity 83.7%; Score 499.4; DB 2; Length 153386;
Matches 536; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```

QY 1 ATGCCGGAACAGAGTAAAGATTACCGCTGTGTGTCGGCGGGCGGCGGTGGCAAG 60
Db 83570 ATGCCGGAACAGAGTAAAGATTACCGCTGTGTGTCGGCGGGCGGCGGTGGCAAG 83511
QY 61 AGCTGCGTGGTGGCTGTGTGTAAGGGCACGTTCCGACACCTTACATCCCAACCATC 120
Db 83510 AGCTGCGTGGTGGCTGTGTGTAAGGGCACGTTCCGACACCTTACATCCCAACCATC 83451
QY 121 GAGGACACCTTCCGGGAGGTATGACGTGCGACAAGCGGTGACAGCTCCAGATCA 180
Db 83450 GAGGACACCTTCCGGGAGGTATGACGTGCGACAAGCGGTGACAGCTCCAGATCA 83391
QY 181 GACACACACCGGACGACCACTTCCGGGACGTGACAGCGCTGTCCATCTCCAAAGGGCCAC 240

```

```

|||||
Db 83390 GACACACACCGGACGACCACTTCCGGGACGTGACAGCGGTGTCCATCTCCAAAGGGCCAC 83331
QY 241 GCCTTCATCCGTGGTCTCTCCGTCACGACGAGTGGTGGAGGAGGCGCCATC 300
Db 83330 GCCTTCATCCGTGGTCTCTCCGTCACGACGAGTGGTGGAGGAGGCGCCATC 83271
QY 301 TACAAGCTCATCTGTCAGATCAAGGAGGCGAGCTGAGAGCATCCCGTGTATCTCTGTGGGC 360
Db 83270 TACAAGCTCATCTGTCAGATCAAGGAGGCGAGCTGAGAGCATCCCGTGTATCTCTGTGGGC 83211
QY 361 AACAACTGCTGATGACGACGCGGAGGTGACACGCGGACGCGAGCGGTGGCCCAA 420
Db 83210 AACAACTGCTGATGACGACGCGGAGGTGACACGCGGACGCGAGCGGTGGCCCAA 83151
QY 421 GAGTGAAGTGGCTTTCATGATGAGACCTCGGCGCAAGATGACTCAACGTCAAGAGCTC 480
Db 83150 GAGTGAAGTGGCTTTCATGATGAGACCTCGGCGCAAGATGACTCAACGTCAAGAGCTC 83091
QY 481 TTCCAGAGCTGCTGACGCTGAGACGCGCGGAAACATGACCTCAACATGACGCGCAAG 540
Db 83090 TTCCAGAGCTGCTGACGCTGAGACGCGCGGAAACATGACCTCAACATGACGCGCAAG 83031
QY 541 CGCTCCGGGAAGCAGAAAGAGGACAGACCGCGCTCAAGGGCAATGCAACCTCATGTGA 597
Db 83030 CGCTCCAGCAGCAAGAAAGAGGCGCGACCGCATCAAGGGCAATGCGGCTCATGTGA 82974

```

RESULT 13
AB062937
LOCUS
DEFINITION
Macaca fascicularis brain cDNA clone:qf1A-13004, full insert
sequence.
ACCESSION
AB062937.1 GI:14388335
VERSION
oligo capping; fls (full insert sequence).
KEYWORDS
Macaca fascicularis adult male frontal lobe left cDNA to mRNA,
clone, lib:macaque brain cDNA library qf1A clone:qf1A-13004.
SOURCE
Macaca fascicularis

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Macaca.
REFERENCE
Osada,N., Hida,M., Kusuda,J., Tanuma,R., Iseki,K., Hirai,M.,
Terao,K., Suzuki,Y., Sugano,S. and Hashimoto,K.
Isolation of full-length cDNA clones from macaque brain cDNA
libraries

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLES
JOURNAL
Direct Submission
Submitted (11-JUN-2001) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail:khashi@nih.go.jp, URL:http://www.nih.go.jp/yoken/genebank/
Tel:81-3-5285-1111(ex.2120), Fax:81-3-5285-1181)
Lab host:
TOP10
Vector:
pME18S-FU3 (ACC.No. AB009864)
R. Site1:
DraIII (CACTGTGTG)
R. Site2:
DraIII (CAACATGTG)
Description: 1st strand cDNA was primed with an oligo(dt) primer
[ATGTCGCTTTTATTTTATTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME18S-FU3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by oligo-capping method
(Sugano et al., Institute of Medical Science, University of
Tokyo).
Custom primer used for sequencing
(5' end primer [CTTCTGCTCAAAAGCTGGC];
3' end primer [CGACCTCAGCTCGACACA]).

COMMENT

FEATURES
source
Location/Qualifiers
1..4412
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="Qf1A-13004"
/sex="male"
/tissue.type="frontal lobe left"
/clone_lib="macaque brain cdna library Qf1A"
/dev_stage="adult"
407..1006
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB60730.1"
/db_xref="GI:14389336"
/translation="MPEQSDNRVAVFGAGVGSLSVLRVYKGFRESYIPTVEDTY
KQVISCDSICTLQITDTTSHQFPMQRLSISKHAFILVTSRSLLEKPYE
QICEIKGDESIPIMLVGNKCDSESPREVOSSEALATWCAFMETSAKLNHNKYE
LFQELINLEKRTVSLQIDGKSKOOKRREKLKRCVIM"
BASE COUNT 1153 a 994 c 1065 g 1200 t
ORIGIN
Query Match 64.6%; Score 385.6; DB 9; Length 4412;
Best Local Similarity 78.8%; Pred. No. 5.4e-50;
Matches 473; Conservative 0; Mismatches 124; Indels 3; Gaps 1;
QY 1 ATGCCGGACAGATACGATTACCGCGTGTGTTCGGGGCGGGCGGCGGCGGCGAG 60
DB 407 ATGCCGAGACAGAGCAACGATTACCGGGTGGCGGTGGGGCGGCGGCGGCGAG 466
QY 61 AGCTCGCTGTGCTGCGCTTCGTGAAGGGGACGCTCCGACACCTACATCCACACAT 120
DB 467 AGCTCCCTGTGTGAGGTTGTGAAGGACACATCCGGAGAGCTACATCCGACGCTG 526
QY 121 GAGGACACTACCGGAGGTGATCAGCTGGCAGACAAGAGCGTGTGACAGCTGACATCACA 180
DB 527 GAAGACACCTACCGGAGGTGATCAGCTGGCAGACAAGAGCATATGACGTTGACAGATCAC 586
QY 181 GACACACCGGAGGACGACGATCCGGGCGATGACGCGCTGTCATCTTCGAAGGGCCAC 240
DB 587 GACACACGGGAGGACGACGATCCGGGCGATGACGCGGTGTCATCTTCGAAGGGCCAC 646
QY 241 GCCTTCATCCTGTGTTCTCCGTACACGACCAAGCAAGTCCGCGAGAGCTGGGGCCCATC 300
DB 647 GCCTTCATCCTGTGTTCTCCGTACACGACGCGGACATCCCTGGAGAGCTCAAGCCCATC 706
QY 301 TCAAGCTCATGTCGATCAAGGAGCGTGGAGGACATCCCGTGTGATGCTGTGGGC 360
DB 707 TACGAACAATCTGCGAGATCAAGGGGAGCATGAGAGCATCCCATCATGCTGTGGGC 766
QY 361 AACAGTGGATGAGACGCGAG---CGGAGGTGGANACCGCGGAGCGGCGAGGGGTGGCC 417
DB 767 AACAGTGGATGAGACGCGGAGCGCGGAGGTGACAGACGAGCGAGGGCGCTTGGGC 826
QY 418 CAAGAGTGAAGTGGCTTTCATGAGAGCTCGGCCAAGATGATGATACAGCTCAAGAG 477
DB 827 CCACATGGAAGTGGCTTTCATGAGAGCTCGGCCAAGCTCAACATTAACGTGAAGAG 886
QY 478 CTCTTCAGAGAGCTGCTGAGAGCGCGCGGAGACATGAGCTCAACATCGAGCGC 537
DB 887 CTCTTCAGAGAGCTGCTGAGAGCGCGGAGACATGAGCTCAACATCGAGCGG 946
QY 538 AAGCGTCCGGAGAGAGAGAGAGACGCGGCTCAAGGGCAATGACCTCATGTGA 597
DB 947 AAAAAAGACAGACAG 1006

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT
MGC.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1875)
Srausberg, R.
Direct Submission
Submitted (21-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: DCTP/DRP
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
Contact: amadansystemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Series: IRAL Plate: 3 Row: P Column: 6.
Location/Qualifiers
1..1875
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:2364 IMAGE:3029902"
/tissue_type="ovary, adenocarcinoma"
/clone_lib="NIH-MGC_9"
/lab_host="DH10B-R"
/note="vector: pORB7"
80..679
/codon_start=1
/product="Similar to CG8500 gene product"
/protein_id="AAH08065.1"
/db_xref="GI:14165525"
/translation="MPEQSDNRVAVFGAGVGSLSVLRVYKGFRESYIPTVEDTY
KQVISCDSICTLQITDTTSHQFPMQRLSISKHAFILVTSRSLLEKPYE
QICEIKGDESIPIMLVGNKCDSESPREVOSSEALATWCAFMETSAKLNHNKYE
LFQELINLEKRTVSLQIDGKSKOOKRREKLKRCVIM"
BASE COUNT 518 a 438 c 472 g 447 t
ORIGIN
Query Match 63.8%; Score 380.8; DB 9; Length 1875;
Best Local Similarity 78.3%; Pred. No. 3.4e-49;
Matches 470; Conservative 0; Mismatches 127; Indels 3; Gaps 1;
QY 1 ATGCCGGACAGATACGATTACCGCGTGTGTTCGGGGCGGGCGGCGGCGGCGAG 60
DB 80 ATGCCGAGACAGATACGATTACCGGGTGGCGGTGGGGCGGCGGCGGCGGCGAG 139
QY 61 AGCTCGCTGTGCTGCGCTTCGTGAAGGGGACGTTCCGAGACACCTACATCCACCATC 120
DB 140 AGCTCCCTGTGTGAGGTTGTGAAGGACACATTCGCGGAGAGCTACATCCGAGCGTG 199
QY 121 GAGGACACTACCGGAGGTGATCAGCTCGACAGAGAGCGTGTGACAGCTGACATCACA 180
DB 200 GAAGACACCTACCGGAGGTGATCAGCTCGACAGAGAGATGACATTCGAGATCAC 259
QY 181 GACACACCGGAGGACGACGATTCGCGGCGATGACGCGCTGTCATCTCCAAAGGGCCAC 240
DB 260 GACACAGCGGGAGGACGACGATTCGCGGCGATGACGCGGCTGTCATCTCCAAAGGGCCAC 319
QY 241 GCCTTCATCCTGTGTTCTCCGTACACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
DB 320 GCCTTCATCCTGTGTTCTCCGTACACGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 379
QY 301 TCAAGCTCATGTCGATCAAGGAGGACGCGTGGAGAGACATCCCGTGTGATGCTGTGGGC 360

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:56 ; Search time 72.4223 Seconds
(without alignments)
8438.824 Million cell updates/sec

Title: US-09-873-546-6

Perfect score: 21

Sequence: 1 gcgtggcgaagaactcgtcg 21

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_rnd:*
37: em_htg_vtl:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19.4	92.4	597	9 AY056037	AY056037 Homo sapi
2	19.4	92.4	597	9 AY059641	AY059641 Homo sapi
3	19.4	92.4	597	10 AY059640	AY059640 Mus muscu
4	19.4	92.4	774	9 AB076888	AB076888 Homo sapi
5	19.4	92.4	1249	6 AX430295	AX430295 Sequence
6	19.4	92.4	1574	10 AB008423	AB008423 Rattus no
7	19.4	92.4	1624	10 RACGYPB2	M16655 Rat cytochr
8	19.4	92.4	1698	10 RACGYPB2	M22330 Rat cytochr
9	19.4	92.4	2763	1 CCU56552	U56652 Caulobacter
10	19.4	92.4	2827	9 AK096600	AK096600 Homo sapi
11	19.4	92.4	3061	6 AX430418	AX430418 Sequence
12	19.4	92.4	3391	9 BC030660	BC030660 Homo sapi
13	19.4	92.4	6710	1 CAJ10321	CAJ10321 Caulobact
14	19.4	92.4	7701	10 RS1ID26	X52027 Rattus norv
15	19.4	92.4	10960	1 AE005869	AE005869 Caulobact
16	19.4	92.4	126901	2 AC091518	AC091518 Mus muscu
17	19.4	92.4	176340	2 AC120750	AC120750 Rattus no
18	19.4	92.4	177540	9 AC006538	AC006538 Homo sapi
19	19.4	92.4	180544	2 AC107527	AC107527 Rattus no
20	19.4	92.4	201320	2 AC073816	AC073816 Mus muscu
21	18.4	87.6	633	1 AF180424	AF180424 Helicobac
22	18.4	87.6	757	1 AF315501	AF315501 Helicobac
23	17.8	84.8	278	1 AF320123	AF320123 Mycobacte
24	17.8	84.8	1034	3 AF255672	AF255672 Drosophil
25	17.8	84.8	1218	9 HSP450	X16865 Human mRNA
26	17.8	84.8	1449	6 AX411044	AX411044 Sequence
27	17.8	84.8	1449	9 HSDBI	X07618 Human mRNA
28	17.8	84.8	1450	6 AX192411	AX192411 Sequence
29	17.8	84.8	1494	6 E10647	E10647 Human cDNA
30	17.8	84.8	1494	6 E10867	E10867 cDNA encodi
31	17.8	84.8	1494	6 E10868	E10868 cDNA encodi
32	17.8	84.8	1494	6 E10869	E10869 cDNA encodi
33	17.8	84.8	1494	6 E10870	E10870 cDNA encodi
34	17.8	84.8	1494	6 MFU38218	U38218 Macaca fasc
35	17.8	84.8	1545	6 E15820	E15820 cDNA encodi
36	17.8	84.8	1563	9 HSP450IID	X16867 Human mRNA
37	17.8	84.8	1566	6 AR084365	AR084365 Sequence
38	17.8	84.8	1566	6 AR084374	AR084374 Sequence
39	17.8	84.8	1567	9 HSDBI	X08006 Homo sapien
40	17.8	84.8	1567	9 HOMCYPDB1	M20403 Human cyto
41	17.8	84.8	1568	6 A20907	A20907 debriisoquin
42	17.8	84.8	1568	6 AR084372	AR084372 Sequence
43	17.8	84.8	1569	6 AR084375	AR084375 Sequence
44	17.8	84.8	1571	6 AR084373	AR084373 Sequence
45	17.8	84.8	1593	6 AX195173	AX195173 Sequence

ALIGNMENTS

RESULT 1
LOCUS AY056037 597 bp mRNA linear PRI 24-JUL-2002
DEFINITION Homo sapiens Rtg protein mRNA, complete cds.
ACCESSION AY056037
VERSION AY056037.1 GI:16555333
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 597)
AUTHORS Ellis,C.A., Vos,M.D., Howell,H., Vallecorse,T., Fuls,D.W. and
Clark,G.J.
TITLE Rtg is a novel Ras-related protein and potential neural tumor

JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

Suppressor
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 9876-9881 (2002)
12107278
2 (bases 1 to 597)
Castro, A.F. and Quilliam, L.A.
Identification of Rlg, a novel Ras family member sharing homology
with NOEY2 and Rap
Unpublished
3 (bases 1 to 597)
Castro, A.F. and Quilliam, L.A.
Direct Submission
Submitted (13-SEP-2001) Biochemistry and Molecular Biology, Indiana
University School of Medicine, 635 Barnhill Drive, MS-4053,
Indianapolis, IN 46202, USA
Location/Qualifiers
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13.3"
/tissue_type="brain; anaplastic oligodendroglioma"
/note="derived from dbEST A1497811"
1..597
/note="Ras family GTPase"
/codon_start=1
/product="Rlg protein"
/protein_id="AAL23715.1"
/db_xref="GI:16535334"
/translation="MPEQSNDRVAVFGAGVGKSSLYLRFVKGTFPRDTYIPTIEDTY
ROVISCDKSVCTLOITDTTGSHQFPAMORLSISKHAFILFVSYSKOSLELGPYIK
LIVQIKGSVEDIPVILVGNKCDETQREVDTREAOQVVAQEMKCAFMTSAKKNYVKEL
FOELLTLETRRMSLNIDGKRSGKRRDRVKGKCTLM"

BASE COUNT 134 a 178 c 195 g 90 t

ORIGIN

Query Match 92.4%; Score 19.4; DB 9; Length 597;
Best Local Similarity 95.2%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
|||||
Db 50 GCGTGGCAAGACTGCTGG 70

RESULT 2
AY059641 597 bp mRNA linear PRI 29-OCT-2001
LOCUS Homo sapiens small GTP-binding tumor suppressor 1 mRNA, complete
DEFINITION cds.
ACCESSION AY059641
VERSION AY059641.1 GI:16508175
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 597)
Gong, L. and Wu, K.
Molecular cloning of GBRSL, a novel gene encoding a small
GTP-binding tumor suppressor
Unpublished
2 (bases 1 to 597)
Gong, L.
Direct Submission
Submitted (12-OCT-2001) Dept. of Cardiology, U.T. MD. Anderson
Cancer Center, 1515 Holcombe Blvd-449, Houston, TX 77030, USA
Location/Qualifiers
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
1..597
/note="GBRSL"

CDS

FEATURES
source

1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
1..597
/note="GBRSL"

/codon_start=1
/product="small gtp-binding tumor suppressor 1"
/protein_id="AAL17968.1"
/db_xref="GI:16508176"
/translation="MPEQSNDRVAVFGAGVGKSSLYLRFVKGTFPRDTYIPTIEDTY
ROVISCDKSVCTLOITDTTGSHQFPAMORLSISKHAFILFVSYSKOSLELGPYIK
LIVQIKGSVEDIPVILVGNKCDETQREVDTREAOQVVAQEMKCAFMTSAKKNYVKEL
FOELLTLETRRMSLNIDGKRSGKRRDRVKGKCTLM"

BASE COUNT 134 a 178 c 195 g 90 t

ORIGIN

Query Match 92.4%; Score 19.4; DB 9; Length 597;
Best Local Similarity 95.2%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
|||||
Db 50 GCGTGGCAAGACTGCTGG 70

RESULT 3
AY059640 597 bp mRNA linear ROD 29-OCT-2001
LOCUS Mus musculus small GTP-binding tumor suppressor 1 (gbts1) mRNA,
complete cds.
DEFINITION complete cds.
ACCESSION AY059640
VERSION AY059640.1 GI:16508173
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 597)
Gong, L. and Wu, K.
Molecular cloning of GBRSL, a novel gene encoding a small
GTP-binding tumor suppressor
Unpublished
2 (bases 1 to 597)
Gong, L.
Direct Submission
Submitted (12-OCT-2001) Dept. of Cardiology, U.T. MD. Anderson
Cancer Center, 1515 Holcombe Blvd-449, Houston, TX 77030, USA
Location/Qualifiers
1..597
/organism="Mus musculus"
/strain="129/SvJ"
/db_xref="taxon:10090"
1..597
/gene="gbts1"
/gene="gbts1"
/note="GBRSL"
/codon_start=1
/product="small GTP-binding tumor suppressor 1"
/protein_id="AAL17967.1"
/db_xref="GI:16508174"
/translation="MPEQSNDRVAVFGAGVGKSSLYLRFVKGTFPRDTYIPTIEDTY
ROVISCDKSVCTLOITDTTGSHQFPAMORLSISKHAFILFVSYSKOSLELGPYIK
LIVQIKGSVEDIPVILVGNKCDETQREVDTREAOQVVAQEMKCAFMTSAKKNYVKEL
FOELLTLETRRMSLNIDGKRSGKRRDRVKGKCTLM"

BASE COUNT 138 a 166 c 194 g 99 t

ORIGIN

Query Match 92.4%; Score 19.4; DB 10; Length 597;
Best Local Similarity 95.2%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
|||||
Db 50 GCGTGGCAAGACTGCTGG 70

RESULT 4

AB076888 774 bp mRNA linear PRI 25-JUN-2002
 LOCUS Homo sapiens mRNA for Di-Ras1, complete cds.
 DEFINITION AB076888
 ACCESSION AB076888
 VERSION AB076888.1 GI:21624247
 KEYWORDS Homo sapiens adult brain cDNA to mRNA, clone_11b:lambda ZAPII human brain cDNA.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Kontani,K., Tada,M., Ogawa,T., Okai,T. and Katada,T.
 TITLE Di-Ras: A distinct Subgroup of Ras-family GTPases with Unique Biochemical Properties
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 774)
 AUTHORS Kontani,K., Ogawa,T., Okai,T., Tada,M. and Katada,T.
 TITLE Direct Submission
 JOURNAL Submitted (24-DEC-2001) Toshiaki Katada, University of Tokyo, Department of Physiological Chemistry, Graduate School of Pharmaceutical Sciences, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan (E-mail:katadaemol.f.u-tokyo.ac.jp, Tel:81-3-5841-4750, Fax:81-3-5841-4751)
 FEATURES
 source 1..774
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="19"
 /map="19p13.3"
 /tissue_type="brain"
 /clone_lib="lambda ZAPII human brain cDNA"
 /dev_stage="adult"
 1..774
 /gene="Di-ras1"
 /protein_id="BAC01115.1"
 /db_xref="GI:21624248"
 /translation="MPEQNDYRVVVFAGVGKSLVLRFGKTFRDYIPTIEDTY ROYISCKSVCTLOITDPTGSHOPAPMORLSISGHFLIFVSTYSOSLEELPIK LYOIRKGSVDIPYMLGVNKCDEPQREVDREARAVNOEMKCAFMEISAKNNYVKEL FOELITETRRNMSLNTDGRSGKOKRTDRYKCKTLA"
 BASE COUNT 154 a 268 c 239 g 113 t
 ORIGIN
 Query Match 92.4%; Score 19.4; DB 9; Length 774;
 Best Local Similarity 95.2%; Pred. No. 53;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GCGTGGGCAAGACTCGCTGG 21
 |||||||
 Db 149 GCGTGGGCAAGAGCTCGCTGG 169
 RESULT 5
 LOCUS AX430295 1249 bp DNA linear PAT 28-JUN-2002
 DEFINITION AX430295
 ACCESSION AX430295
 VERSION AX430295.1 GI:21655659
 KEYWORDS human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.
 TITLE Molecules for disease detection and treatment

JOURNAL Patent: WO 0240715-A 77 23-MAY-2002;
 INCYTE GENOMICS INC (US)
 FEATURES
 source 1..1249
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /note="Incyte ID No: LG:389870.1:2000SEP08"
 BASE COUNT 230 a 436 c 355 g 228 t
 ORIGIN
 Query Match 92.4%; Score 19.4; DB 6; Length 1249;
 Best Local Similarity 95.2%; Pred. No. 54;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GCGTGGGCAAGACTCGCTGG 21
 |||||||
 Db 100 GCGTGGGCAAGAGCTCGCTGG 120
 RESULT 6
 LOCUS AB008423 1574 bp mRNA linear ROD 29-OCT-1997
 DEFINITION Rattus norvegicus mRNA for CYP2D2, complete cds.
 ACCESSION AB008423
 VERSION AB008423.1 GI:2575858
 KEYWORDS CYP2D2.
 SOURCE Rattus norvegicus (strain:Sprague-Dawley) 7 weeks age male liver
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (sites)
 AUTHORS Wan,J., Imaoka,S., Chow,T., Hiroi,T., Yabusaki,Y. and Funae,Y.
 TITLE Expression of four rat CYP2D isoforms in Saccharomyces cerevisiae and their catalytic specificity
 JOURNAL Arch. Biochem. Biophys. (1997) In press
 AUTHORS 2 (bases 1 to 1574)
 Imaoka,S.
 TITLE Direct Submission
 JOURNAL Submitted (22-OCT-1997) Susumu Imaoka, Osaka City University Medical School, Laboratory of Chemistry, 1-4-54 Asahimachi,, Abeno-ku, Osaka 545, Japan (E-mail:Imaoka@msc.med.osaka-cu.ac.jp, Tel:81-6-645-2081, Fax:81-6-646-3922)
 FEATURES
 source 1..1574
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /sex="male"
 /tissue_type="liver"
 /dev_stage="7 weeks age"
 /tissue_lib="cDNA"
 1..1574
 /gene="Cyp2d2"
 /note="Cyp2d2"
 /gene="Cyp2d2"
 /note="Cyp2d2 cytochrome P450 isoform"
 /codon_start=1
 /product="CYP2D2"
 /protein_id="BAA23123.1"
 /db_xref="GI:2575859"
 /translation="MGILIGDDIANAVYIFTAIFLLINDVYHRHKEFWTAHPPGPVPLP GLGNDLOVDFENMPYSLYKLSRGGVDFSLDIANKPVPVYINGLAIVRLVYTGEDTA DRPLPIYNIHGYGNKSKGVVALPAGPWRQRRSVSTLDFGVGKSLQWVTEEA GHLDCTFAKEAHEHPNPSILSKAVSNVIAVLVAREVEDPFRNRKLTLSKSTFEA DTGMAEVILNALIPILQIPGLPGKVFPLNSFIALVDMLIEHKKSMDPAPRSDMTD AFLAEMOKAKNPSPSSDENLRLVVIDLFMAAGVYTTSTSMALLIILHDPQRYV HEETIDVYGVRRPEMDAORPFTNAYIHFVORPADIVPNTIHMTSRODKRGFLI PKGTTILPNLSAVYKDETVMKPLRFHREHHLVDAGNFKVKEAHMPRSAGRACLGEF LARMELEFETCLQRFSESVLAGRPSTHGVALPVPQYDLCAYAR"
 BASE COUNT 343 a 456 c 416 g 359 t
 ORIGIN

Query Match 92.4%; Score 19.4; DB 10; Length 1574;
Best Local Similarity 95.2%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTGGCAGAACTCGCTGG 21
|||||
Db 437 GCGTGGCAGAAATCGCTGG 457

RESULT 7
RATCYPDB2
LOCUS Rat cytochrome P450-db2 mRNA, complete cds. linear ROD 27-Apr-1993
DEFINITION
ACCESSION M16655
KEYWORDS M16655.1 GI:203835
SOURCE cytochrome P450; debrisoquine 4-hydroxylase.
Rat (Sprague-Dawley, 8 week old) liver microsomes, cDNA to mRNA,
clone pP450db2.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1624)
Gonzalez, F.J., Matsunaga, T., Nagata, K., Meyer, U.A., Nebert, D.W.,
Pastewka, J., Kozak, C.A., Gillette, J., Gelboin, H.V. and
Hardick, J.P.
Debrisoquine 4-hydroxylase: characterization of a new P450 gene
subfamily, regulation, chromosomal mapping, and molecular analysis
of the DA rat polymorphism
DNA 6 (2), 149-161 (1987)
JOURNAL MEDLINE 87217961
PUBMED 3582092

FEATURES
source
location/Qualifiers
1..1624
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
<1..1624
/product="CYP450 mRNA"
51..1553
/note="debrisoquine 4-hydroxylase."
/codon_start=1
/protein_id="AA41055.1"
/translation="MGILIGDDLMNAVITAFLLVLDLVHRRKFTAHYPPGPVPLP
GCGNLQVDFENMPYSLYKLRSGYGVSLQIAKRPVVVINGLAKARELLVYGEDTA
DRPLPTVNHILGYGDKSGVLAAPGPERDRRSVTLRPGVGRKSLQDMVTEEA
GHLCDFEAKAEHPNPSTLLSKAVSNVAVARFEYEDPFENRMKLTKESEGE
DTGFMAEVNATPILQITGLPGKVPKLNSTFALVDKMLIEHKKSMDAPDPDMTD
AFLEAOKAKGNPESFENDENLRVVIDLFMGWVTSTLLSMALLMLHDPQORV
HEIDEVIGQVRPEKADQARMPLTNVATHEVQRFADIVPTNIPHTSRIKIFQGLI
PKGTTLIPNLSSVLDKETWEKPLRHPHFLDQGNFVKHAEFMSAGRACJGEP
LARMELFLFTCLORFSVLAGRRPSTHGVALPVPPOYOLCAVAR"

BASE COUNT 347 a 466 c 437 g 374 t

ORIGIN Unreported.

Query Match 92.4%; Score 19.4; DB 10; Length 1624;
Best Local Similarity 95.2%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTGGCAGAACTCGCTGG 21
|||||
Db 487 GCGTGGCAGAAATCGCTGG 507

RESULT 8
RATCYPDB2
LOCUS Rat cytochrome P450 mRNA, complete cds. linear ROD 27-Apr-1993
DEFINITION
ACCESSION M22330
KEYWORDS M22330.1 GI:203823
SOURCE cytochrome P450.
Rat (Sprague-Dawley, male) liver, cDNA to mRNA, clone

ORGANISM
lambda-circMF[22.1].
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1698)
Ishida, N., Tawaragi, Y., Inuzuka, C., Sugita, O., Kubota, I.,
Nakazato, H., Noguchi, T. and Sassa, S.
Four species of cDNAs for cytochrome P450 isozymes immunorelated to
P450C-M/E encode for members of P450IId subfamily, increasing the
number of members within the subfamily
Biochem. Biophys. Res. Commun. 156 (2), 681-688 (1988)
JOURNAL MEDLINE 89050091
PUBMED 3190674

COMMENT Draft entry and computer-readable sequence for [1] kindly provided
by N. Ishida, 20-JAN-1989.

FEATURES
source
location/Qualifiers
1..1698
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
<1..1698
/product="CYP mRNA"
125..1627
/note="cytochrome P450"
/codon_start=1
/protein_id="AA41049.1"
/db_xref="GI:203824"
/translation="MGILIGDDLMNAVITAFLLVLDLVHRRKFTAHYPPGPVPLP
GCGNLQVDFENMPYSLYKLRSGYGVSLQIAKRPVVVINGLAKARELLVYGEDTA
DRPLPTVNHILGYGDKSGVLAAPGPERDRRSVTLRPGVGRKSLQDMVTEEA
GHLCDFEAKAEHPNPSTLLSKAVSNVAVARFEYEDPFENRMKLTKESEGE
DTGFMAEVNATPILQITGLPGKVPKLNSTFALVDKMLIEHKKSMDAPDPDMTD
AFLEAOKAKGNPESFENDENLRVVIDLFMGWVTSTLLSMALLMLHDPQORV
HEIDEVIGQVRPEKADQARMPLTNVATHEVQRFADIVPTNIPHTSRIKIFQGLI
PKGTTLIPNLSSVLDKETWEKPLRHPHFLDQGNFVKHAEFMSAGRACJGEP
LARMELFLFTCLORFSVLAGRRPSTHGVALPVPPOYOLCAVAR"

BASE COUNT 373 a 483 c 450 g 392 t

ORIGIN 311 bp upstream of HindIII site.

Query Match 92.4%; Score 19.4; DB 10; Length 1698;
Best Local Similarity 95.2%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTGGCAGAACTCGCTGG 21
|||||
Db 561 GCGTGGCAGAAATCGCTGG 581

RESULT 9
CCU56652
LOCUS CCU56652 2763 bp DNA linear BCT 14-NOV-1996
DEFINITION Caulobacter crescentus lon protease (lon) gene, complete cds.
ACCESSION U56652
KEYWORDS U56652.1 GI:1336062

SOURCE
ORGANISM Caulobacter crescentus.
Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
Caulobacter.
1 (bases 1 to 2763)
Wright, R., Stephens, C., Zweiger, G., Shapiro, L. and Alley, M.R.
Caulobacter lon protease has a critical role in cell-cycle control
of DNA methylation
Genes Dev. 10 (12), 1532-1542 (1996)
JOURNAL MEDLINE 8666236
PUBMED 8666236

REFERENCE
AUTHORS 2 (bases 1 to 2763)
Wright, R.J., Stephens, C., Zweiger, G., Shapiro, L. and Alley, M.R.K.
TITLE Direct Submission
JOURNAL Submitted (24-Apr-1996) Developmental Biology, Stanford University,
Beckman Center B300, Stanford Medical Center, Stanford, CA
94305-5427, USA
location/Qualifiers


```

source
1. .2763
/organism="Caulobacter crescentus"
/strain="NAL1000"
/db_xref="taxon:155892"
340. .2740
/gene="lon"
340. .2739
/gene="lon"
/codon_start=1
/transl_table=11
/product="lon protease"
/protein_id="AAB18765.1"
/db_xref="GI:1667399"
/translacion="MSELRLLPLVLRDIYVPPHVVPLFVGRDSEVPMRGDK
OILVLTOKNSADDDPAPDIFRGVIAVLIOLLPCTVIVLGGKRAAVSPTDG
ESTYEAQIGEVSEDDGAPPEAGLSRAVEDPENVYKANKVPEPALSTIPIAEPK
LADSIKRAHLSVIGDKONLEIFDVYKLEVFVLMSEISLVLOEKIKRVRKOME
KTQREYLYNEQMAKIQRELIDPDARDELIDLERIKKTKLSKEARTVASELKLKN
MSPMSAEVTVRNLYLDWLSIPMGKATKKIDLVSEGIILADHYGLEKVERILEYL
AVOARTNSLKGPILCLVPGPGVTSLSKSTAKVAREFVMSIGGVGVAEIRGHR
TYISMGKVYOSMKKAKTNAFVLIDEDIDKSGDYRGPASALLEVDPSQNSFCD
HYLEVDYELSVMEVTTANSLNMPOLIDRMEIRIPICTEDEKLEIKRILRLAK
DHGLKPAEFIVPDKAIRDLRYVTRAGVRSLEBELGALKATYDRLAREVASTID
DERLAKYAGVKKRYRGTEVDQVGIYVGLMTEFGDILITIEAVKMGKGRMOITGN
LKDYAKESIAANSYVRSALQFGIKPVEFKTQVHIHPDGATPKDPSAGIAMALA
MVSVLTGIPRKDIAMGETILRGRTVAILGLKEKLLAALRSVKTVLIPQENEDLA
DYPTGVKDGLEIIFVSTYDEVKLHNLGRLTPVEMNEAEPIITTSAKKDDSDAMLT
H"

BASE COUNT      564 a      817 c      923 g      459 t
ORIGIN

Query Match
Best Local Similarity 95.2%; Score 19.4; DB 1; Length 2763;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTCGCTGG 21
|||||
Db 1403 GCGTGGCAAGACTCGCTGG 1423

RESULT 10
AK096600      2827 bp mRNA linear PRI 15-JUL-2002
LOCUS Homo sapiens cDNA FLJ39281 fls, clone OCBBF2011067, highly similar
to Homo sapiens chingulin mRNA.
ACCESSION AK096600.1 GI:21756131
VERSION AK096600.1
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens fetal brain cDNA to mRNA, clone_lib:OCBBF2
clone:OCBBF2011067.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuna,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masubo,Y., Nagai,K. and Isogai,T.
NEBO human cDNA sequencing project
Unpublished
2 (bases 1 to 2827)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomiceshri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEBO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan

```

```

FEATURES
source
1. .2827
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="OCBBF2011067"
/risue_type="brain"
/clone_lib="OCBBF2"
/dev_stage="fetal"
/note="Cloning vector: pME18SFL3"

BASE COUNT      688 a      750 c      983 g      406 t
ORIGIN

Query Match
Best Local Similarity 95.2%; Score 19.4; DB 9; Length 2827;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTCGCTGG 21
|||||
Db 179 GCGTGGCAAGACTCGCTGG 199

RESULT 11
AX430418      3061 bp DNA linear PAT 28-JUN-2002
LOCUS AX430418
DEFINITION Sequence 200 from Patent WO240715.
ACCESSION AX430418
VERSION AX430418.1 GI:21655782
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
Chalup,M.S., Altus,C.M., Lincoln,S.F., Dufour,G.E. and Jackson,S.
Molecules for disease detection and treatment
Patent: WO 0240715-A 200 23-MAY-2002;
INCYTE GENOMICS INC (US)
FEATURES
source
1. .3061
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: LI:332161.1:2000SEP08"

BASE COUNT      606 a      840 c      927 g      688 t
ORIGIN

Query Match
Best Local Similarity 92.4%; Score 19.4; DB 6; Length 3061;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTCGCTGG 21
|||||
Db 100 GCGTGGCAAGACTCGCTGG 120

RESULT 12
BC030660      3391 bp mRNA linear PRI 21-MAY-2002
LOCUS BC030660
DEFINITION Homo sapiens, similar to R1g protein, clone MGC:33391
IMAGE:4814337, mRNA, complete cds.
ACCESSION BC030660
VERSION BC030660.1 GI:21040534
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3391)
Strausberg,R.
Direct Submission

```

JOURNAL

Submitted (20-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbioology.org>
Contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Keltman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRK Plate: 46 Row: e Column: 8
This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction.

FEATURES

source

Location/Qualifiers
1..3391
/organism="Homo sapiens"
/db_xref="LocusID:148252"
/db_xref="taxon:9606"
/clone="MGC:33391 IMAGE:4814337"
/tissue_type="Brain, hippocampus"
/clone_1ib="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
138..734
/codon_start=1
/product="similar to Rlg protein"
/protein_id="AAH30660.1"
/db_xref="GI:21040535"
/translation="MEQSNDRVYVFGAGVGKSLVLPVKGFRRDYPITIEDTY ROVISCDKSVCTLIQITDTGSHQFPAMORLISKGAFILVFSVSKOSELGPIYK LIVOIGKSVIEDIPVMLVGNKCDTQREVDTRAOAVAOEMKCAFMTSKMNYVKEL FQELTLLETRRMSNLINDGKRSKQKRTDRVAKGKCTLM"

CDS

BASE COUNT 650 a 996 c 1016 g 729 t
ORIGIN

Query Match

Best Local Similarity 92.4%; Score 19.4; DB 9; Length 3391;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GCCTGGCGAAGACTGCTGG 21
|||||

Db 187 GCCTGGCGAAGACTGCTGG 207

RESULT 13

LOCUS

DEFINITION

CAJ10321 6710 bp DNA linear BCT 10-MAR-2001
Caulobacter crescentus partial tlg gene and clpX, cica, clpX, lon genes.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AJ010321
AJ010321.1 GI:3688203
cica gene; clp gene; ClpX gene; endopeptidase Clp; lon gene; tlg gene; trigger factor.
Caulobacter crescentus.
Caulobacter crescentus
Bacteria; Proteobacteria; alpha subdivision; Caulobacter group; Caulobacter.

REFERENCE

AUTHORS

TITLE

1 (bases 1 to 6710)
Fuchs,T., Wiget,P., Osteras,M. and Jenal,U.
Precise amounts of a member of a phosphotransferase superfamily are essential for growth and normal morphology in Caulobacter crescentus
Mol. Microbiol. 39 (3), 679-692 (2001)

PUBMED

11169108
2 (bases 1 to 6710)

REFERENCE

AUTHORS

TITLE

JOURNAL

PUBLISHED

SOURCE

FEATURES

source

Submitted (18-AUG-1998) Jenal U., University of Basel, Biozentrum, Klingelbergstrasse 70, Basel, CH-4056, SWITZERLAND

gene

1..1065
/gene="tlg"
/gene="tlg"
/function="peptidylprolyl isomerase"
/codon_start=1

CDS

/transl_table=11
/product="trigger factor"
/protein_id="CA09089.1"
/db_xref="GI:3688204"
/db_xref="SWISS-PROT:O87705"
/translation="EPEPIDPTSIELVKPVYKVSDEVOEALDELAQARTYEPRTGK SIKARDGDLIDFVGTIDGVEFAGKAGAEVLVSGQFTGFEDQILGARGDDVY VVKRPEEYQAKDLAKQDAEFATKYQEVNAPADGKADDELRIGISDLAATLILKS NLAGRTDSSRFKLRALDLVDTHDFPLPRWDAAEFAGIWOQVEADKARGLPPE DAKEETDQLDKEYKRIKIERVRILGLVLAIGKNDVYVDDELTAIMEARQYGAQA QGVFDMYRQRODLQAALRAPIDYEDKVDLIFGKAKIEKEVSKDELLERBDLP EGYGC"

terminator

gene

promoter

RBS

CDS

terminator

gene

CDS

gene

promoter

intron	4205..4304 /gene="P450 IID2" /number=8 4305..>4483 /gene="P450 IID2" /number=9 /function="codes for a region of P450 protein that is associated with the noncovalently bound heme iron at the enzymes active site."
exon	
polyA_signal	4533..4538 /gene="P450 IID2" 4634
polyA_site	
BASE COUNT	2039 a 1783 c 2019 g 1859 t 1 others
ORIGIN	
Query Match	92.4%; Score 19.4; DB 10; Length 7701;
Best Local Similarity	95.2%; Pred No. 57;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0	
1 GCCTGGCGACAACTCGCTGG 21	
GCCTGGCGACAAGAAATCGCTGG 1894	

RESULT 15					
AE005869/c					
LOCUS	AE005869	10960 bp	DNA	linear	BCT 12-JUN-2002
DEFINITION	Caulobacter crescentus CB15 section 195 of 359 of the complete genome.				
ACCESSION	AE005869	AE005673			
VERSION	AE005869.1	GI:13423417			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE	1 (bases 1 to 10960)				
AUTHORS	Nierman,W.C., Feldblum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R., Pococha,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., DeBoy,R.T., Dodson,R.J., Dukkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khouri,H., Shetty,J.M., Berry,K., Uterback,T., Tran,K., Wolf,A., Vamathavan,J., Ermolaev,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and Fraser,C.M.				
TITLE	Complete genome sequence of Caulobacter crescentus				
JOURNLT.	Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)				
PUBMED	11259647				
REFERENCE	2 (bases 1 to 10960)				
AUTHORS	Nierman,W.C., Feldblum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.				

Nelson, W.C., Newton, A., Stephens, C., Pladde, N.D., Ely, B.,
 Laub, M.T., DeRoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L.,
 Haft, D.H., Kolonay, J.F., Smic, J., Craven, M., Khouri, H., Shetty, J.,
 Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J.,
 Ermolaeva, M., White, O., Salzberg, S.L., Shapiro, L., Venter, J.C. and
 Fraser, C.M.
 Direct Submission
 Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1..10960
 /organism="Caulobacter crescentus CB15"
 /strain="CB15"
 /db_xref="taxon:190650"
 131..207
 /gene="tRNA-Asp-1"
 131..207
 /gene="tRNA-Asp-1"
 272..348
 /product="tRNA-Asp"
 272..348

```

gene
CDS
    /gene="tRNA-Asp-2"
    272. .348
    /gene="tRNA-Asp-2"
    /product="tRNA-Asp"
    407. .1420
    /gene="CC1957"
    407. .1420
    /gene="CC1957"
    /note="identified by Glimmer2; putative"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="AAK23932.1"
    /db_xref="GI:13423418"
    /translation="MRHLADMTSALPLPTISVLDLAPVPGGSSVQALKNISIDLAQHAH
    TGLHRLVLAIEHNMPGLIASATFAVVICQILAAATTTIVGSGVLMPPHAPLMAVEQ
    GLFLNALPQRIDILGIRAPGTQATMRALRYRAGVDSFADVDVELQIMFPAASDQ
    VRAVPEAGGVADVITWIGSSITWGAQLAALGLPYAFAAHFAPDALIEALTLYRHHKPP
    ESIADRPYAGGIVGCAADTDEAKRLSTSTQOFLALRRGRGLLPVVDITREHASI
    GLAGLDHTLGYAAITGSPETVARKIDRVLELTLGDELMAAAQIHDHARRRSYELAA
    LRA"
    complement(1417. .2625)
    /gene="CC1958"
    complement(1417. .2625)

```

/note="identified by match to protein family HMM"
 /codon_start=1
 /transl_table=1
 /product="glycosyl transferase, group 1 family protein"
 /protein_id="AAK23933.1"
 /db_xref="GI:13423419"
 /translation="MTDTRPACSLGPGLOPETRSLGDEPRVLVDTMTLYAPRSGGVR
 YLSSRAMLASHRPVRRHTLVPGARDSDHDSGGVSIYAALPFGAGYRMPVYRSGV
 ERLIQGDGLLEAGDPYRPPGLAALRAGCALPYVPGFCHTDLKLAHLIHIDMAEKPK
 OKRMAAITYRQEDQAVASPRFTAGRLTEGVDAIGLPGLVGTALFHPERGRDGLRRR
 LGLSKDLILVAPGAPAREKKLDVLAVERLDGPYKLLFVAGAGGAVSDRTLCIDV
 VRDPSKDLASCDAPFHANDNEPFGILVEANACGLPVVAGSAGVAESVDAAVGG
 LAVASEATPAFADALELPERDVTVGEAARRRAVERHGMADVTELCLITYLRGRRR
 FSSAHHGH"
 /complement(2928..3002)
 /gene="tRNA-Val-1"
 /complement(2928..3002)
 /gene="tRNA-Val-1"
 /product="tRNA-Val-1"
 /complement(3109..3387)
 /gene="CC1959"
 /complement(3109..3387)
 /gene="CC1959"
 /note="identified by match to protein family HMM"
 /codon_start=1
 /transl_table=1
 /product="DNA-binding protein HU"
 /protein_id="AAK23934.1"
 /db_xref="GI:13423420"
 /translation="MTTKAEELVTAIAEKAGINKNQANDALEAFLEAVTSLKSGDVR
 LVGCTFEKAVTRAGTARNPRTGCTVNRPAKTRAPQVGEGLKSSLS"
 /complement(3531..5930)
 /gene="CC1960"
 /complement(3531..5930)
 /gene="CC1960"
 /note="identified by match to TIGR protein family HMM
 TIGR00763"
 /codon_start=1
 /transl_table=1
 /product="ATP-dependent protease LA"
 /protein_id="AAK23935.1"
 /db_xref="GI:13423421"
 /translation="MSELRTLPVLPRLDIVFPHMVVPFLFVGDRKSVRLAEVARGDK
 QILVLTQKNSADDDPARGDIFEVGLVATVLIQLKPGTVKRVESGARAAVASFQD
 ESYEATQKNSDDPARGDAPRAEASRAVVDQFENVYKLNKKVPPEALASIPQIEPKR
 LADSIAHLKRTIDKONLEIFDVNRLEKVRALMGEELSVLQVEKRTISRAKQKE
 KTQEVYINDEMKALIDRELSDPPDARDELIDLEKRIKTKLSKAPRTAKSELKLETL
 MSPSASIVTNRNALIDRELSDPPDARDELIDLEKRIKTKLSKAPRTAKSELKLETL

AVOARTNSLKPLICLVGPPVGKTSLSKSIATAKGRFVRMSIGVDEAIEIRGHR
TYIGSDPGKVVQSMKAKTNNATVLDIEDIKMGSDVGRDPAASALLEVDLSDNSSTED
HYLEVDYDSQVWMEFTVTANSLNMPDPLDMETIRIDPGTEDEKELIAKHITLPIKLAK
DHGLKRAEPIVPPKAIKIDLRITRYTREAGVNSLERELGALKARTVRLAKRKSASITID
DEBLAKYAGVKKTRYGETDEVDQVIGVTLAMEFEQDILITIAVKRPGKGRMOITGN
LKDVMEKESIANSYVSRALQGIKPPVEKEDVNIHVDSGATPKPGSPAGIAMALA
MVSVLGIPIRKIDIAMTGETILGRVTAIGLEKEKLIALAASGVKTYVLIQOENKDLA
DVPGTVKGLLEIIPSVTVEDEVLKHALTGPLTPVEMWNAEPIITTSAKKDDGSDAMLT
H"
/complement(6325..7587)
/gene="CC1961"
/complement(6325..7587)
/gene="CC1961"
/note="Identified by match to TIGR protein family HMM
TIGR00382"
/codon_start=1
/transl_table=11
/product="ATP-dependent Clp protease, ATP-binding subunit
ClpX"
/protein_id="AAK23936.1"
/db_xref="GI:13423422"
/translation="MTKRAAGDTRKSTLYCSFCKSQHEVRKLIAGPTVFCDECEVLC
MDLIREBKIAFVKSQDVPPTPREICEVDYVIGQGHAKVLAVAHNHKRLNHS
KNDVDELAKSNILLVGPTGKTLAQTLARIIDVPTMADATLTLAGYVGEDVENI
VLKLDADYNVERAQRGIYIDEIDKISRKSNSPSTIRDSGEGVQALAKIMEGTIV
ASVPGGRKHPOOEFLQVDTNILFCGGAFAGLEKIIISARGAASIGFAKTVQPE
ERRTGETLRNVEPDDLOREGLIPEPTLPVATIEDLDEALVLIKILTPPKNAFVKQY
QRLPEMNIIGLTETEDALHQAATKARTKARGSRISIMEGILLETMFELPTIEGVEE
VVVNAEYVEGRAPLTIYAEKKGGAASA"
7971..8591
/gene="CC1962"
7971..8591
/gene="CC1962"
/note="Identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="cIca protein"
/protein_id="AAK23937.1"
/db_xref="GI:13423423"
/translation="MTGEMGHEPLVAFDFDGTLYVKSFSNAPLKMWRAGRWGRVLR
LTPALLAYVDRNRKIKKAAYROFLAKGATVAQIENDARPAEAFAPSLRPDAVAVM
RGRARAKAKVIVTASPDLTVAFPARGLADLIGTLRCSDDRIIGLDGNNCRAK
EKVIRLEVEFGPVRLTAAAGDYSGDEMLAIADKGYRIFRGKPA"
complement(8670..9302)
/gene="CC1963"
complement(8670..9302)
/gene="CC1963"
/note="Identified by match to TIGR protein family HMM
TIGR00493"
/codon_start=1
/transl_table=11
/product="ATP-dependent Clp protease, proteolytic subunit"
/protein_id="AAK23938.1"
/db_xref="GI:13423424"
/translation="MMYDPVSTAMNLVPMVVEQTSRGERAFDIFSRLKRIIFLTGP
VEDGMSLICAQLFLESENPKKEIAMYINSPGVYAGLAIDTMQIKSPVSTVCM
GMAASMSLLAAGAQRISLPMARIMVQPSGGFRGQASDIERHAEDIIKTKRRLN
EIVYKHGGRTEVEERTLDRDHMSADEAKAMGIVDHVYSRDAEAGAE"
complement(9480..10838)
/gene="CC1964"
complement(9480..10838)
/gene="CC1964"
/note="Identified by match to TIGR protein family HMM
TIGR00115"
/codon_start=1
/transl_table=11
/product="trigger factor"
/protein_id="AAK23939.1"
/db_xref="GI:13423425"
/translation="MOIVERKSGSLSVFGVTVPASRLATRLIARIAEVAPOMVVKG
RPGKVPAAHRYKRALMGEVIEQALNETTYVLEDNKLRPACQPELNPSSDMDKYI
AGGEDLSFDLAVEMPEFEPIDPTSTELVAVTKVSDSEVQEADELAKOARTPEPT
GKSLAKKDGQLLIDFVGTIDGVEFAGKAELVYSGQFI PGFEDQLVGAKPGDD

BASE COUNT 1854 a 3619 c 3396 g 2091 t
ORIGIN

Query Match 92.4%; Score 19.4; DB 1; Length 10960;
Best Local Similarity 95.2%; Pred. No. 57;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GCGTGGCGAAGACTCGCTGG 21
|||||
Db 4867 GCGTGGCGAAGACTCGCTGG 4847

Search completed: March 23, 2003, 16:10:00
Job time : 77.4223 secs

VVVKFPEEYQAKDLAKDAEFATKVOEVRAPVDGKADELAKRLGLSDLAUTELL
KSNLAGRYDSSRRPKRLRALLDVLDTHDPLPRMVDADBPAGIMQGVADKARGLP
PDAEKTEFDLKDQEKRIARVRRLSLVLAIEIGKNNVYVTDLTPAIRARQYGA
QAGQVFMTRQRADLQALRAPTYEDKVVDLIFEKATIEEKVSKDELLEDLPEST
GG"

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:37 ; Search time 24.4179 Seconds
(without alignments)
1936.775 Million cell updates/sec

Title: US-09-873-546-6

Perfect score: 21

Sequence: 1 gcgtggcgaagaactgcgtg 21

Scoring table:

Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_101002.*
1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19.4	92.4	1249	24	ABO72525	Human MDR1 encodin
2	19.4	92.4	3061	24	ABO72648	Human MDR1 encodin
3	17.8	84.8	130	24	ABK87227	Partial human cyto
4	17.8	84.8	400	24	AAK40743	DNA sequence of am
5	17.8	84.8	477	19	AAV19498	Cytochrome P45011d
6	17.8	84.8	702	23	ABL14159	Drosophila melanog
7	17.8	84.8	1449	24	ABN97193	Gene #3691 used to
8	17.8	84.8	1450	22	AAD09849	Human CYP2D6 gene.
9	17.8	84.8	1494	16	AAO87729	Human auxiliary cy

10	17.8	84.8	1494	16	AAO87730	Human auxiliary cy
11	17.8	84.8	1494	16	AAO87731	Human auxiliary cy
12	17.8	84.8	1494	16	AAO87732	Human auxiliary cy
13	17.8	84.8	1494	17	AAT28395	Human cytochrome p
14	17.8	84.8	1494	17	AAT28396	Human cytochrome p
15	17.8	84.8	1494	17	AAT28397	Human cytochrome p
16	17.8	84.8	1494	17	AAT28398	Human cytochrome p
17	17.8	84.8	1494	17	AAT17388	Human derived cyto
18	17.8	84.8	1494	24	ABO72216	Human CYP2D6 gene
19	17.8	84.8	1545	19	AAV19496	Cytochrome P45011d
20	17.8	84.8	1586	12	AAQ12893	Debrisoquine hydro
21	17.8	84.8	1593	22	AAD09937	Human drug metabol
22	17.8	84.8	2951	23	ABL14158	Drosophila melanog
23	17.8	84.8	3493	23	ABL141094	Drosophila melanog
24	17.8	84.8	6472	24	ABO72215	Human CYP2D6 gene
25	17.8	84.8	6472	24	ABO72364	Human CYP2D6 gene
26	17.8	84.8	9432	24	ABD34213	Human cytochrome p
27	17.4	82.9	922	23	ABL04759	Drosophila melanog
28	17.4	82.9	1471	23	ABL15253	Drosophila melanog
29	17.4	82.9	3350	23	ABL04758	Drosophila melanog
30	17.4	82.9	4389	23	ABL04744	Drosophila melanog
31	17.4	82.9	14112	23	ABL15252	Drosophila melanog
32	16.8	80.0	421	23	AAS82963	DNA encoding novel
33	16.8	80.0	657	24	ABK14232	Human related RAS
34	16.8	80.0	1021	24	AB199215	Mouse ischaemic co
35	16.8	80.0	1078	23	AAS82965	DNA encoding novel
36	16.8	80.0	1188	23	ABL11393	Drosophila melanog
37	16.8	80.0	1629	24	ABO9083	M. capsulatus gene
38	16.8	80.0	2223	22	AAH41606	Human R-Ras, c-Raf
39	16.8	80.0	2223	24	ABL50193	Human R-Ras, c-Raf
40	16.8	80.0	2223	24	ABL50225	Human R-Ras, c-Raf
41	16.8	80.0	2236	17	AAT08525	Oncogene R-Ras mut
42	16.8	80.0	2336	24	ABN95766	Gene #2264 used to
43	16.8	80.0	2813	23	ABL26308	Drosophila melanog
44	16.8	80.0	3188	23	ABL11392	Drosophila melanog
45	16.8	80.0	5686	24	ABK14231	Human related RAS

ALIGNMENTS

RESULT 1	
ABO72525	ABO72525 standard; CDNA: 1249 BP.
AC	ABO72525;
DT	03-SEP-2002 (first entry)
DE	Human MDR1 encoding CDNA SEQ ID NO 77.
XX	Human: MDR1; disease detection and treatment molecule polynucleotide;
KW	proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
KW	autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
KW	rheumatoid arthritis; transgenic; gene therapy; antiatherosclerotic;
KW	hepatotropic; antiinflammatory; antiposoriatic; cytostatic; anti-HIV;
KW	antiallergic; antianemic; antiasthmatic; antiatherosclerotic; antiout;
KW	neuroprotective; antirheumatic; antiarthritic; gene; ss.
OS	Homo sapiens.
XX	
PN	WO200240715-A2.
XX	
PD	23-MAY-2002.
PF	06-SEP-2001; 2001WO-US27628.
XX	
PR	06-SEP-2000; 2000US-230505P.
XX	
PR	06-SEP-2000; 2000US-230514P.
XX	
PR	06-SEP-2000; 2000US-230515P.
XX	
PR	06-SEP-2000; 2000US-230517P.
XX	
PR	06-SEP-2000; 2000US-230518P.
XX	
PR	06-SEP-2000; 2000US-230519P.

PR 06-SEP-2000: 2000US-230595P.
PR 06-SEP-2000: 2000US-230597P.
PR 06-SEP-2000: 2000US-230598P.
PR 06-SEP-2000: 2000US-230599P.
PR 06-SEP-2000: 2000US-230610P.
PR 06-SEP-2000: 2000US-230655P.
PR 06-SEP-2000: 2000US-230988P.
PR 06-SEP-2000: 2000US-230989P.
PR 07-SEP-2000: 2000US-230951P.
PR 07-SEP-2000: 2000US-231163P.
PR 07-SEP-2000: 2000US-231167P.
PA (INCYTE GENOMICS INC.
PI Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
PI Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
PI Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
PI Gerstlin EH, Peralta CH, David MH, Panzer SR, Flores V, Dafio A;
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
XX MPI: 2002-527544/56.
XX P-PSDB: ABP51307.
XX
XX Novel human disease detection and treatment polypeptide, useful in
XX diagnosis, prevention or treatment of cell proliferative disorders e.g.
XX arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
XX e.g. AIDS
XX
XX Claim 1: Page 350-351, 618pp: English.
XX
XX The invention relates to an isolated human disease detection and
XX treatment (MDPT) polypeptide (I) selected from a polypeptide having a
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the
XX specification, a naturally occurring polypeptide comprising a sequence
XX having at least 90% identity to (I) or a biologically active or
XX immunogenic fragment of (I). (I) is useful for screening a compound for
XX effectiveness as an agonist or antagonist, for screening a compound that
XX specifically binds (I) or modulates the activity of (I), and for
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.
XX Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
XX screening a compound for effectiveness in altering expression of a target
XX polynucleotide comprising. Oligonucleotides and antibodies are useful for
XX detecting MDPT in a sample or for assessing toxicity of a test compound,
XX in a diagnostic test for a condition or a disease associated with the
XX expression of MDPT in a biological sample, for detecting (I) in a sample,
XX and for purifying (I) from a sample. A composition comprising (I), an
XX agonist or antagonist is useful for treating a disease or condition
XX associated with decreased or increased expression of functional MDPT.
XX (I) or (II) are useful for diagnosing, treating or preventing disorders
XX associated with aberrant expression of MDPT, where the disorders are
XX selected from a cell proliferative disorder such as arteriosclerosis,
XX cirrhosis, hepatitis, psoriasis, and cancer and an
XX autoimmune/inflammatory disorder such as AIDS, Addison's disease,
XX allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or
XX rheumatoid arthritis. (II) are useful for creating knockin humanised
XX animals or transgenic animals to model human diseases, in somatic or
XX germine gene therapy, to generate a transcript image of a tissue or cell
XX type, for detecting differences in the chromosomal location due to
XX translocation or inversion among normal, carrier or affected individuals
XX and as hybridisation probes for mapping naturally occurring genomic
XX sequences.
XX
XX Sequence 1249 BP; 230 A; 436 C; 355 G; 228 T; 0 other;
SQ
XX
XX Query Match 92.4%; Score 19.4; DB 24; Length 1249;
XX Best Local Similarity 95.2%; Pred. No. 4.4;
XX Matches 20; Conservativity 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 GCGTGGGCAAGACTGCGTGG 21
XX ||||||||| |||||||
DB 100 GCGTGGGCAAGACTGCGTGG 120

RESULT 2
ABQ72648
ID ABQ72648 standard; cDNA; 3061 BP.
XX
XX AC ABQ72648;
XX
XX 03-SEP-2002 (first entry)
XX
XX Human MDPT encoding cDNA SEQ ID NO 200.
DE
XX
XX Human; MDPT; disease detection and treatment molecule polynucleotide;
XX proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
XX autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
XX rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerosis;
XX hepatocytic; antiinflammatory; antiproliferative; cytostatic; anti-HIV;
XX antiangiogenic; antianemic; antistatic; antihypertensive; antitumor;
XX neuroprotective; antineurotic; antineurotic; gene; ss.
XX
XX Homo sapiens.
XX
XX WO200240715-A2.
XX
XX 23-MAY-2002.
XX
XX 06-SEP-2001: 2001WO-US27628.
XX
XX 06-SEP-2000: 2000US-230505P.
XX 06-SEP-2000: 2000US-230514P.
XX 06-SEP-2000: 2000US-230515P.
XX 06-SEP-2000: 2000US-230517P.
XX 06-SEP-2000: 2000US-230518P.
XX 06-SEP-2000: 2000US-230519P.
XX 06-SEP-2000: 2000US-230595P.
XX 06-SEP-2000: 2000US-230597P.
XX 06-SEP-2000: 2000US-230598P.
XX 06-SEP-2000: 2000US-230599P.
XX 06-SEP-2000: 2000US-230610P.
XX 06-SEP-2000: 2000US-230655P.
XX 06-SEP-2000: 2000US-230988P.
XX 06-SEP-2000: 2000US-230989P.
XX 07-SEP-2000: 2000US-230951P.
XX 07-SEP-2000: 2000US-231163P.
XX 07-SEP-2000: 2000US-231167P.
XX
XX (INCYTE GENOMICS INC.
XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;
XX Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE, Dahl CR;
XX Momiyama MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM;
XX Gerstlin EH, Peralta CH, David MH, Panzer SR, Flores V, Dafio A;
XX Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;
XX MPI: 2002-527544/56.
XX P-PSDB: ABP51431.
XX
XX Novel human disease detection and treatment polypeptide, useful in
XX diagnosis, prevention or treatment of cell proliferative disorders e.g.
XX arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder
XX e.g. AIDS
XX
XX Claim 1: Page 414; 618pp: English.
XX
XX The invention relates to an isolated human disease detection and
XX treatment (MDPT) polypeptide (I) selected from a polypeptide having a
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the
XX specification, a naturally occurring polypeptide comprising a sequence
XX having at least 90% identity to (I) or a biologically active or
XX immunogenic fragment of (I). (I) is useful for screening a compound for
XX effectiveness as an agonist or antagonist, for screening a compound that
XX specifically binds (I) or modulates the activity of (I), and for
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.
XX Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for
XX screening a compound for effectiveness in altering expression of a target

CC polynucleotide comprising, oligonucleotides and antibodies are useful for
CC detecting MDR in a sample or for assessing toxicity of a test compound,
CC in a diagnostic test for a condition or a disease associated with the
CC expression of MDR in a biological sample, for detecting (I) in a sample,
CC and for purifying (I) from a sample. A composition comprising (I), an
CC agonist or antagonist is useful for treating a disease or condition
CC associated with decreased or increased expression of functional MDR.
CC (I) or (II) are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of MDR, where the disorders are
CC selected from a cell proliferative disorder such as arteriosclerosis,
CC cirrhosis, hepatitis, psoriasis, and cancer and an
CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,
CC allergy, anemia, asthma, atherosclerosis, gout, multiple sclerosis or
CC rheumatoid arthritis. (II) are useful for creating knockin humanised
CC animals or transgenic animals to model human diseases, in somatic or
CC germline gene therapy, to generate a transcript image of a tissue or cell
CC type, for detecting differences in the chromosomal location due to
CC translocation or inversion among normal, carrier or affected individuals
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences.
XX
SQ Sequence 3061 BP; 606 A; 840 C; 927 G; 688 T; 0 other;
Query Match 92.4%; Score 19.4; DB 24; Length 3061;
Best Local Similarity 95.2%; Pred. No. 4.9;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GCGTGGCAAGACTCGCTGG 21
DB 100 GCGTGGCAAGACTCGCTGG 120
RESULT 3
ABK87227
ID ABK87227 standard; DNA, 130 BP.
XX
AC ABK87227;
XX
DT 24-SEP-2002 (first entry)
XX
DE Partial human cytochrome 2D6 gene sequence.
XX
KW Restriction endonuclease recognition sequence; RERS; drug resistance;
KW human genetic disease; pathogenic microorganism; bladder carcinoma;
KW sickle-cell anaemia; thalassemia; Alzheimer's disease; phenylketonuria;
KW galactosemia; Wilson's disease; diabetes insipidus; neurofibromatosis;
KW familial hypercholesterolemia; genetic analysis; hereditary disease;
KW tumour diagnosis; disease predisposition; forensic; paternity;
KW crop cultivation; animal breeding; expression profiling;
KW infectious organism; plant; food safety; cytochrome 2D6;
KW single nucleotide polymorphism; SNP; gene; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT variation replace (65..67,CG)
FT /*tag= a
FT /standard_name= "Single nucleotide polymorphism"
FT /note= "SNP"
XX
PN W020024647-A2.
PD 13-JUN-2002.
XX
PF 01-OCT-2001; 2001WO-US30742.
XX
PR 02-OCT-2000; 2000US-237409P.
PR 10-NOV-2000; 2000US-247166P.
PR 10-NOV-2000; 2000US-247167P.
PR 10-NOV-2000; 2000US-247172P.
PR 10-NOV-2000; 2000US-247173P.
PR 10-NOV-2000; 2000US-247275P.
PR 24-JAN-2001; 2001US-263971P.
PR

PR 15-FEB-2001; 2001US-269244P.
PR 21-JUN-2001; 2001US-300319P.
PR 21-JUN-2001; 2001US-300350P.
PR 27-JUN-2001; 2001US-301394P.
XX
PA (KECK-) KECK GRADUATE INST.
XX
PI Van Ness J, Galas DJ, Garrison LK;
XX
DR WPI: 2002-527924/56.
XX
PT Identifying single nucleotide polymorphisms at defined positions in
PT target nucleic acids, by utilising oligonucleotide primers that contain
PT a part of an interrupted restriction endonuclease recognition sequence
XX
PS Example: Page 54; 82pp; English.
XX
CC The present invention relates to a new method of identifying a nucleotide
CC at a defined position in a single stranded target nucleic acid. The
CC method of the invention utilises a pair of oligonucleotide primers
CC comprising a nucleotide sequence that is complementary to a nucleotide
CC sequence of the target nucleic acid or complementary to a nucleotide
CC sequence of the complement of the target nucleic acid and further
CC comprising part of an interrupted restriction endonuclease recognition
CC sequence (RERS). The invention is useful for identifying a nucleotide at
CC a defined position in a single-stranded target nucleic acid. The defined
CC position may be polymorphic or associated with a disease, especially a
CC human genetic disease, or drug resistance of a pathogenic microorganism.
CC The defined position is associated with a disease, including a human
CC genetic disease e.g. bladder carcinoma, sickle-cell anaemia,
CC thalassemias, Alzheimer's disease, phenylketonuria, galactosemia, or
CC Wilson's disease, diabetes insipidus, familial hypercholesterolemia, or
CC neurofibromatosis. The invention finds applications in genetic analysis
CC for hereditary disease, tumour diagnosis, disease predisposition,
CC forensics or paternity, crop cultivation and animal breeding, expression
CC profiling of cell function and/or disease marker genes, and
CC identification and/or characterisation of infectious organisms that
CC cause infectious disease in plants or animals and/or that are related
CC to food safety. The method is useful for determining gene variations in
CC T-cell receptor genes encoding variable, antigen-specific regions that
CC are involved in the recognition of various foreign antigens. The present
CC nucleic acid sequence represents the partial human cytochrome 2D6 gene
CC sequence that was used in the method of the invention.
XX
SQ Sequence 130 BP; 19 A; 40 C; 46 G; 25 T; 0 other;
Query Match 84.8%; Score 17.8; DB 24; Length 130;
Best Local Similarity 90.5%; Pred. No. 22;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGTGGCAAGACTCGCTGG 21
DB 40 GCGTGGCAAGACTCGCTGG 60
RESULT 4
AAL40743
ID AAL40743 standard; DNA, 400 BP.
XX
AC AAL40743;
XX
DT 25-SEP-2002 (first entry)
XX
DE DNA sequence of amplifier containing CYP450-2D6-G1749C SNP.
XX
KW Variation site: analysing; point mutation; detecting pathogen; SNP;
KW single nucleotide polymorphisms; paternity dispute; prenatal testing;
KW forensic analysis; CYP450-2D6-G1749C; ds.
OS unidentified.
XX
PN W0200194546-A2.
PN

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 702 BP; 185 A; 187 C; 220 G; 110 T; 0 other;

Query Match 84.8%; Score 17.8; DB 23; Length 702;

Best Local Similarity 90.5%; Pred. No. 26;
 Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

YY 1 GCCTGGCAGAACTCGCTGG 21

DB 83 GCCTGGCAGAACTCGCTGG 103

RESULT 7

ID ABN97193 standard; DNA: 1449 BP.

AC ABN97193;

DT 13-AUG-2002 (first entry)

DE Gene #3691 used to diagnose liver cancer.

KM Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;

KM metastatic liver tumor; cytosolic; expression profile; disease state;

KM disease progression; drug toxicity; drug efficacy; drug metabolism.

OS Homo sapiens.

PN W0200229103-A2.

PD 11-APR-2002.

PF 02-OCT-2001; 2001WO-US30589.

PR 02-OCT-2000; 2000US-237054P.

PA (GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;

DR WPI; 2002-426119/45.

PT Diagnosing and detecting the progression of liver cancer,

PT hepatocellular carcinoma or metastatic liver tumor in a patient,

PT involves detecting the level of expression of two or more genes in a

PT liver tissue sample -

PS Claim 1; SEQ ID NO 3691; 298bp; English.

XX The invention relates to a novel method for diagnosing and detecting the

XX progression of liver cancer, hepatocellular carcinoma or metastatic liver

XX tumor in a patient, and differentiating metastatic liver cancer from

XX hepatocellular carcinoma in a patient, involving detecting the level of

XX expression of two or more genes represented in ABN93503-ABN97455 in a

XX tissue sample. The method of the invention has hepatotropic, and

XX cytosolic activity. The method is useful for diagnosing and detecting

XX the progression of liver cancer, hepatocellular carcinoma and metastatic

XX liver carcinoma in a patient. The method is useful for identifying

XX expression profiles which serve as useful diagnostic markers as well as

XX markers that can be used to monitor disease states, disease progression,

XX drug toxicity, drug efficacy and drug metabolism.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1449 BP; 256 A; 465 C; 439 G; 289 T; 0 other;

Query Match 84.8%; Score 17.8; DB 24; Length 1449;

Best Local Similarity 90.5%; Pred. No. 29;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

YY 1 GCCTGGCAGAACTCGCTGG 21

DB 131 GCCTGGCAGAACTCGCTGG 151

RESULT 8

ID AAD09849 standard; DNA: 1450 BP.

AC AAD09849;

DT 12-SEP-2001 (first entry)

DE Human CYP2D6 gene.

KM Polymorphism; amplification; CYP2D6; cytochrome P450; CYP; human;

KM drug metabolism; psychiatric disorder; cardiovascular disorder; ds.

OS Homo sapiens.

PN W0200149883-A2.

PD 12-JUL-2001.

PF 22-DEC-2000; 2000WO-US35186.

PR 30-DEC-1999; 99US-0173699.

PA (ABBO) ABBOTT LAB.

PI Katz DA, Gentile-Davey MC, Cornwell MJ, Huff JB;

DR WPI; 2001-441898/47.

PT Detecting a mutation in target nucleic acid sequence in test sample, by

PT amplifying target and standard nucleic acid sequence using primers,

PT hybridizing probes to the products to form hybrids, and detecting

PT hybrids -

PS Example 1; Page 30; 35pp; English.

XX The invention relates to a method for detecting polymorphism in a target

XX nucleic acid sequence using amplification technique. The method involves

XX amplifying the target sequence and a standard nucleic acid sequence

XX using primers to form amplification products, hybridizing a first

XX labelled probe to the target sequence amplification product and a

XX second labelled probe to the standard sequence amplification product,

XX detecting the signals from the first and the second probe, and comparing

XX the signals to determine the polymorphism. The method is useful for

XX detecting polymorphism in various nucleic acid sequences e.g. CYP2D6 gene

XX which is a member of cytochrome P450 (CYP) gene family. CYP2D6 plays a

XX role in the metabolism of several drugs, including those used for

XX treating psychiatric and cardiovascular disorders. Polymorphism in

XX the CYP2D6 gene have varying effect on an individual's ability to

XX metabolise drugs. The method is suitable for detecting amplification

XX products from multiple and different types of polymorphisms on a single

XX automated platform. The present sequence is human CYP2D6 gene.

XX Sequence 1450 BP; 270 A; 395 C; 521 G; 264 T; 0 other;

Query Match 84.8%; Score 17.8; DB 22; Length 1450;

Best Local Similarity 90.5%; Pred. No. 29;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

YY 1 GCCTGGCAGAACTCGCTGG 21

DB 150 GCCTGGCAGAACTCGCTGG 170

RESULT 9

AA087729

ID AA087729 standard; CDNA: 1494 BP.

```
XX AAQ87729;
XX
XX 15-NOV-1995 (first entry)
XX
XX Human auxiliary cytochrome P450 species 2D6 coding region.
XX
XX Human cytochrome P450: amplification; PCR; primer; expression vector;
XX yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
XX carcinogen; mutagen; liver metabolism; ds.
XX
XX Homo sapiens.
XX
XX EP644267-A.
XX
XX 22-MAR-1995.
XX
XX 20-JUL-1994; 94EP-0111298.
XX
XX 21-JUL-1993; 93JP-0180246.
XX 20-JUL-1993; 93JP-0201120.
XX 30-JUL-1993; 93JP-0208279.
XX
XX (HAYASHI) HAYASHI K.
XX (SUMO) SUMITOMO CHEM CO LTD.
XX
XX Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
XX Yabusaki Y;
XX WPI: 1995-116991/16.
XX P-PSDB: AAR72375.
XX
XX Evaluation of safety of a chemical cpd. - using recombinant yeast
XX expressing human cytochrome p450 and a yeast NADPH-P450 reductase
XX
XX Examples; Page 82-84, 124pp; English.
XX
XX The nucleotide sequence of the cDNA coding region for the human
XX auxiliary cytochrome P450 species 2C6. The gene encodes a protein of 497
XX amino acids. The cDNA was amplified by PCR using the primers AAQ87763-6.
XX The product was cloned into the yeast expression vectors pAAH5N or pAHR
XX to produce the vectors p2D6 for the expression of the cytochrome P450
XX alone or p2D6R for co-expression with the yeast NADPH-P450 reductase.
XX The vectors are used in a method for evaluating the safety of a chemical
XX compound by reacting the chemical compound with recombinantly produced
XX human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9 (AAQ87715),
XX 2E1 (AAQ87716), or 3A4 (AAQ87717) or their auxiliary species and
XX variants (AAQ87718-32), and yeast NADPH-P450 reductase, either as a fused
XX protein or in cell extracts, and analysing the resulting metabolite to
XX assess the safety of the chemical compound. The method is useful for
XX determining whether the chemical compound, or its metabolite, will be
XX converted into a carcinogenic or mutagenic form through metabolism in the
XX liver.
XX
XX Sequence 1494 BP; 248 A; 508 C; 446 G; 292 T; 0 other;
XX
XX Query Match 84.8%; Score 17.8; DB 16; Length 1494;
XX Best Local Similarity 90.5%; Pred. No. 29;
XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 GCGTGGCAAGAACTCGCTGG 21
XX ||||||||| |||||
XX Db 428 GCGTGGCAAGAACTCGCTGG 448
XX
XX RESULT 10
XX ID AAQ87730 standard; cDNA: 1494 BP.
XX AC AAQ87730;
XX XX
XX DT 15-NOV-1995 (first entry)
XX XX
```

```
DE Human auxiliary cytochrome P450 species 2D6 variant 1 coding region.
XX
XX Human cytochrome P450: amplification; PCR; primer; expression vector;
XX yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
XX carcinogen; mutagen; liver metabolism; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX variation 886
XX FT /*tag=
XX FT /note= "T to C change in variant 1 changes amino
XX FT acid from Cys to Arg"
XX
XX EP644267-A.
XX
XX 22-MAR-1995.
XX
XX 20-JUL-1994; 94EP-0111298.
XX
XX 21-JUL-1993; 93JP-0180246.
XX 20-JUL-1993; 93JP-0201120.
XX 30-JUL-1993; 93JP-0208279.
XX
XX (HAYASHI) HAYASHI K.
XX (SUMO) SUMITOMO CHEM CO LTD.
XX
XX Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
XX Yabusaki Y;
XX WPI: 1995-116991/16.
XX P-PSDB: AAR72376.
XX
XX Evaluation of safety of a chemical cpd. - using recombinant yeast
XX expressing human cytochrome p450 and a yeast NADPH-P450 reductase
XX
XX Examples; Page 87-89, 124pp; English.
XX
XX The nucleotide sequence of the cDNA coding region for the human
XX auxiliary cytochrome P450 species 2D6 variant 1. The gene contains a
XX change at base 886 from T to C as compared to the wild type sequence
XX (AAQ87729). This changes the amino acid residue 296 from Cys to Arg. The
XX cDNA was amplified by PCR using the primers AAQ87763-6. The product was
XX cloned into the yeast expression vectors pAAH5N or pAHR to produce the
XX vectors p2D6 variant 1 for the expression of the cytochrome P450 alone
XX or p2D6R variant 1 for co-expression with the yeast NADPH-P450
XX reductase. The vectors are used in a method for evaluating the safety of
XX a chemical compound by reacting the chemical compound with recombinantly
XX produced human cytochrome P450 molecular species 1A2 (AAQ87714), 2C9
XX (AAQ87715), 2E1 (AAQ87716), or 3A4 (AAQ87717) or their auxiliary species
XX and variants (AAQ87718-32), and yeast NADPH-P450 reductase, either as a
XX fused protein or in cell extracts, and analysing the resulting metabolite
XX to assess the safety of the chemical compound. The method is useful for
XX determining whether the chemical compound, or its metabolite, will be
XX converted into a carcinogenic or mutagenic form through metabolism in the
XX liver.
XX
XX Sequence 1494 BP; 248 A; 509 C; 446 G; 291 T; 0 other;
XX
XX Query Match 84.8%; Score 17.8; DB 16; Length 1494;
XX Best Local Similarity 90.5%; Pred. No. 29;
XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 GCGTGGCAAGAACTCGCTGG 21
XX ||||||||| |||||
XX Db 428 GCGTGGCAAGAACTCGCTGG 448
XX
XX RESULT 11
XX ID AAQ87731 standard; cDNA: 1494 BP.
XX AC AAQ87731;
XX XX
```

```

XX 15-NOV-1995 (first entry)
XX
XX Human auxillary cytochrome P450 species 2D6 variant 2 coding region.
XX
XX Human auxillary cytochrome P450; amplification; PCR; primer; expression vector;
XX yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
XX carcinogen; mutagen; liver metabolism; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX variation 886
XX /*tag= a
XX /*note= "T to C change in variant 2 changes amino
XX variation 1457
XX /*tag= b
XX /*note= "C to G change in variant 2 changes amino
XX acid from Thr to Ser"
XX
XX EP644267-A.
XX
XX 22-MAR-1995.
XX
XX 20-JUL-1994; 94EP-0111298.
XX
XX 21-JUL-1993; 93JP-0180246.
XX 20-JUL-1993; 93JP-0201120.
XX 30-JUL-1993; 93JP-0208279.
XX
XX (HAYASHI K.
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
XX Yabusaki Y;
XX
XX WPI: 1995-116991/16.
XX P-PSDB: AAR72377.
XX
XX Evaluation of safety of a chemical cpd. - using recombinant yeast
XX expressing human cytochrome p450 and a yeast NADPH-P450 reductase
XX
XX Examples; Page 91-93; 124pp; English.
XX
XX The nucleotide sequence of the cDNA coding region for the human
XX auxillary cytochrome P450 species 2D6 variant 2. The gene contains
XX variations at bases 886: T to C and 1457: C to G as compared to the wild
XX type sequence (AA087729). These change the amino acid residues 296: Cys
XX to Arg and 486: Thr to Ser. The cDNA was amplified by PCR using the
XX primers AA087763-6. The product was cloned into the yeast expression
XX vectors pAMRN or pAMRR to produce the vectors p2D6 variant 2 for the
XX co-expression of the cytochrome P450 alone or p2D6R variant 2 for the
XX co-expression with the yeast NADPH-P450 reductase.
XX The vectors are used in a method for evaluating the safety of a chemical
XX compound by reacting the chemical compound with recombinantly produced
XX human cytochrome P450 molecular species 1A2 (AA087714), 2C9 (AA087715),
XX 2E1 (AA087716), or 3A4 (AA087717) or their auxillary species and
XX variants (AA087718-22), and yeast NADPH-P450 reductase, either as a fused
XX protein or in cell extracts, and analysing the resulting metabolite to
XX assess the safety of the chemical compound. The method is useful for
XX determining whether the chemical compound, or its metabolite, will be
XX converted into a carcinogenic or mutagenic form through metabolism in the
XX liver.
XX
XX Sequence 1494 BP; 248 A; 508 C; 447 G; 291 T; 0 other:
XX
XX Query Match 84.8%; Score 17.8; DB 16; Length 1494;
XX Best Local Similarity 90.5%; Pred. No. 29;
XX Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 GCGTGGCAGACGCTGG 21
XX
XX 11

```

```

DB      428  GCCTGGCCAGAGATCGCTGG  448
RESULT 12
AA087732
ID      AA087732 standard; cDNA; 1494 BP.
XX
AC      AA087732;
XX
DT      15-NOV-1995 (first entry)
XX
DE      Human auxillary cytochrome P450 species 2D6 variant 3 coding region.
XX
KM      Human cytochrome P450: amplification; PCR; primer; expression vector;
KM      yeast NADPH-P450 reductase; safety; fusion protein; metabolite;
KM      carcinogen; mutagen; liver metabolism; ds.
XX
OS      Homo sapiens.
XX
FH      Key
FH      variation
FH      location/Qualifiers
FT      1457
FT      /*tag= a
FT      /note= "C to G change in variant 1 changes amino
XX      acid from Thr to Ser"
XX
PN      EP644267-A.
XX
PD      22-MAR-1995.
XX
PF      20-JUL-1994; 94EP-0111298.
XX
PR      21-JUL-1993; 93JP-0180246.
PR      20-JUL-1993; 93JP-0201120.
PR      30-JUL-1993; 93JP-0208279.
XX
PA      (HAYA/) HAYASHI K.
PA      (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI      Hayashi K, Kaneko H, Komai K, Nakatsuka I, Sakaki T;
PI      Yabusaki Y;
XX
DR      WPI: 1995-116991/16.
DR      P-PSDB; AAR72378.
XX
XX      Evaluation of safety of a chemical cpd. - using recombinant yeast
XX      expressing human cytochrome p450 and a yeast NADPH-P450 reductase
XX
XX      Examples; Page 95-97; 124pp: English.
XX
XX      The nucleotide sequence of the cDNA coding region for the human
XX      auxillary cytochrome P450 species 2D6 variant 3. The gene contains a
XX      change at base 1457 from C to G as compared to the wild type sequence
XX      (AA087729). This changes the amino acid residue 296 from Thr to Ser. The
XX      cDNA was amplified by PCR using the primers AA087763-6. The product was
XX      cloned into the yeast expression vectors pAAH5N or pAHR to produce the
XX      vectors pD6 variant 3 for the expression of the cytochrome P450 alone
XX      or p2DR variant 3 for co-expression with the yeast NADPH-P450
XX      reductase. The vectors are used in a method for evaluating the safety of
XX      a chemical compound by reacting the chemical compound with recombinantly
XX      produced human cytochrome P450 molecular species 1A2 (AA087714). 2C9
XX      (AA087715), 2B1 (AA087716), or 3A4 (AA087717) or their auxillary species
XX      and variants (AA087718-32), and yeast NADPH-P450 reductase, either as a
XX      fused protein or in cell extracts, and analysing the resulting metabolite
XX      to assess the safety of the chemical compound. The method is useful for
XX      determining whether the chemical compound, or its metabolite, will be
XX      converted into a carcinogenic or mutagenic form through metabolism in the
XX      liver.
XX
SQ      Sequence 1494 BP; 248 A; 507 C; 447 G; 292 T; 0 other;
Query Match      84.8%; Score 17.8; DB 16; Length 1494;
Best Local Similarity 90.5%; Pred. No. 29;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 23, 2003, 15:14:42 ; Search time 5.54252 Seconds
(without alignments)
2938.192 Million cell updates/sec

Title: US-09-873-546-6
Perfect score: 21
Sequence: 1 gcgtggcagacgctgg 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 53826 seqs, 387737923 residues
Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PC7_NEM_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PC7US_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	17.8	84.8	1449	10 US-09-880-107-3688	Sequence 3688, Ap
2	17.8	84.8	1450	10 US-09-747-538-1	Sequence 1, Appli
3	17.8	84.8	9432	9 US-09-942-310-1	Sequence 44, Appli
4	16.8	80.0	1157	9 US-10-108-605-44	Sequence 1, Appli
5	16.8	80.0	68750	9 US-10-014-717-1	Sequence 10636, A
6	16.4	78.1	256	10 US-09-960-352-10636	Sequence 2434, Ap
7	16.4	78.1	348	10 US-09-960-352-2434	Sequence 10478, A
8	16.4	78.1	422	10 US-09-960-352-10478	Sequence 13143, A
9	16.2	77.1	385	10 US-09-960-352-13143	Sequence 4246, Ap
10	16.2	77.1	398	10 US-09-983-965-4246	Sequence 4253, Ap
11	16.2	77.1	452	10 US-09-960-352-4253	Sequence 769, App
12	16.2	77.1	454	10 US-09-764-864-769	Sequence 5830, App
13	16.2	77.1	498	10 US-09-960-352-5830	Sequence 354, App
14	16.2	77.1	1443	10 US-09-764-864-354	Sequence 224, App
15	16.2	77.1	1511	9 US-09-962-832-224	Sequence 11, Appli
16	16.2	77.1	1964	10 US-10-078-650-11	Sequence 367, App
17	16.2	77.1	2746	10 US-09-925-297-367	Sequence 1463, Ap
18	16.2	77.1	3641	10 US-09-917-800A-1463	Sequence 1, Appli
19	16.2	77.1	3641	9 US-10-078-650-1	Sequence 1, Appli

20	15.8	75.2	185	10 US-09-294-093B-5564	Sequence 5564, Ap
21	15.8	75.2	295	10 US-09-294-093B-1650	Sequence 1650, Ap
22	15.8	75.2	348	10 US-09-960-352-1110	Sequence 1110, Ap
23	15.8	75.2	1749	9 US-09-809-391-54	Sequence 54, Appli
24	15.8	75.2	8439	10 US-09-764-877-3959	Sequence 3959, Ap
25	15.8	75.2	9875	10 US-09-764-877-3960	Sequence 3960, Ap
26	15.8	75.2	31412	9 US-10-109-551-3	Sequence 3, Appli
27	15.8	75.2	78056	9 US-10-109-551-1	Sequence 1, Appli
28	15.4	73.3	48	10 US-09-880-732-56	Sequence 56, Appli
29	15.4	73.3	47	10 US-09-880-732-55	Sequence 55, Appli
30	15.4	73.3	4059	9 US-09-784-554B-1	Sequence 1, Appli
31	15.2	72.4	473	10 US-09-833-790-372	Sequence 372, App
32	15.2	72.4	570	12 US-10-104-484-1	Sequence 1, Appli
33	15.2	72.4	570	12 US-10-104-484-3	Sequence 3, Appli
34	15.2	72.4	751	10 US-09-770-149-6	Sequence 6, Appli
35	15.2	72.4	1387	10 US-09-925-302-217	Sequence 217, App
36	15.2	72.4	1832	10 US-09-822-849A-407	Sequence 407, App
37	15.2	72.4	2136	9 US-09-938-842A-1480	Sequence 1480, Ap
38	15.2	72.4	2508	9 US-09-738-626-2305	Sequence 2305, Ap
39	15.2	72.4	2760	10 US-09-934-868-51	Sequence 51, Appli
40	15.2	72.4	3750	9 US-09-964-868-32	Sequence 32, Appli
41	15.2	72.4	6254	9 US-09-764-868-1495	Sequence 1495, Ap
42	15.2	72.4	6457	10 US-09-880-107-3389	Sequence 3389, Ap
43	15.2	72.4	6491	10 US-09-954-456-2118	Sequence 2118, Ap
44	14.8	70.5	311	10 US-09-864-761-27685	Sequence 27685, A
45	14.8	70.5	412	10 US-09-960-352-14532	Sequence 14532, A

ALIGNMENTS

RESULT 1
US-09-880-107-3688
Sequence 3688, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880.107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211.379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237.054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3688
LENGTH: 1449
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 X07618
US-09-880-107-3688
Query Match 84.8%; Score 17.8; DB 10; Length 1449;
Best Local Similarity 90.5%; Pred. No. 9.2;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GCGTGGCAGACGCTGG 21
DB 131 GCGTGGCAGACGCTGG 151
RESULT 2
US-09-747-538-1
Sequence 1, Application US/09747538
Patent No. US20020102549A1
GENERAL INFORMATION:
APPLICANT: Abbott Laboratories

```

; APPLICANT: Katz, David A.
; APPLICANT: Gentile-Davey, Maria C.
; APPLICANT: Cornwell, Michael C.
; APPLICANT: Huff, Jeffrey B.
; APPLICANT: Yu, Hong
; TITLE OF INVENTION: AMPLIFICATION BASED POLYMORPHISM
; FILE REFERENCE: 6652.US.01
; CURRENT APPLICATION NUMBER: US/09/747,538
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1450
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-747-538-1

Query Match
Best Local Similarity 84.8%; Score 17.8; DB 10; Length 1450;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTGGGCAAGACTGCTGG 21
Db 150 GCCTGGGCAAGACTGCTGG 170

RESULT 3
US-09-942-310-1
; Sequence 1, Application US/09942310
; Publication No. US20030044797A1
; GENERAL INFORMATION:
; APPLICANT: Ristinger, Carl
; APPLICANT: Andersson, Maria K.
; APPLICANT: Lewander, Tommy
; APPLICANT: Olafsson, Erik
; TITLE OF INVENTION: Detection of CYP2D6 Polymorphisms
; FILE REFERENCE: G6119.1US
; CURRENT APPLICATION NUMBER: US/09/942,310
; CURRENT FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: GB 0021286.0
; PRIOR FILING DATE: 2000-08-30
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 9432
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-942-310-1

Query Match
Best Local Similarity 84.8%; Score 17.8; DB 9; Length 9432;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTGGGCAAGACTGCTGG 21
Db 3300 GCCTGGGCAAGACTGCTGG 3320

RESULT 4
US-10-108-605-44
; Sequence 44, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Broadus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
```

```

; PRIOR APPLICATION NUMBER: US 09/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 1157
; TYPE: DNA
; ORGANISM: Drosophila melanogaster
US-10-108-605-44

Query Match
Best Local Similarity 80.0%; Score 16.8; DB 9; Length 1157;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTGGGCAAGACTGCTG 20
Db 356 GTGTGGGCAAGACTGCTG 375

RESULT 5
US-10-014-717-1
; Sequence 1, Application US/10014717
; Publication No. US20020192778A1
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/10/014,717
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US/09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-10-014-717-1

Query Match
Best Local Similarity 80.0%; Score 16.8; DB 9; Length 68750;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTGGGCAAGACTGCTG 20
Db 21493 GCCTGGGCAAGACTGCTG 21512

RESULT 6
US-09-960-352-10636/C
; Sequence 10636, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 10636
; LENGTH: 256
; TYPE: DNA
```

ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 46-LIB3058-004-Q1-E1-D10
US-09-960-352-10636

Query Match 78.1%; Score 16.4; DB 10; Length 256;
Best Local Similarity 94.4%; Pred. No. 39;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGGCGAAGACTCGCTGG 21
DB 142 TGACGACAGAACTCGCTGG 125

RESULT 7

US-09-960-352-2434
Sequence 2434, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 2434
LENGTH: 348
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 11-LIB3058-014-Q1-K1-C3
US-09-960-352-2434

Query Match 78.1%; Score 16.4; DB 10; Length 348;
Best Local Similarity 94.4%; Pred. No. 40;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGGCGAAGACTCGCTGG 21
DB 203 TGACGACAGAACTCGCTGG 220

RESULT 8

US-09-960-352-10478
Sequence 10478, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 10478
LENGTH: 422
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 45-LIB3058-014-Q1-K1-D2
US-09-960-352-10478

Query Match 78.1%; Score 16.4; DB 10; Length 422;
Best Local Similarity 94.4%; Pred. No. 41;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 TGGCGAAGACTCGCTGG 21
DB 200 TGACGACAGAACTCGCTGG 217

RESULT 9
US-09-960-352-13143

Sequence 13143, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 13143
LENGTH: 385
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 56-LIB3058-047-Q1-K1-F12
US-09-960-352-13143

Query Match 77.1%; Score 16.2; DB 10; Length 385;
Best Local Similarity 85.7%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAGAACTCGCTGG 21
DB 299 GCGTGGCAGAACTCGCTGG 319

RESULT 10

US-09-983-965-4246
Sequence 4246, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983,965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 4246
LENGTH: 398
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
OTHER INFORMATION: Clone ID: 61-LIB3058-040-Q1-K1-H2
US-09-983-965-4246

Query Match 77.1%; Score 16.2; DB 10; Length 398;
Best Local Similarity 85.7%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAGAACTCGCTGG 21
DB 363 GCGTGGCAGAACTCGCTGG 383

RESULT 11

US-09-960-352-4253
Sequence 4253, Application US/09960352
Patent No. US20020137139A1

```
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4253
; LENGTH: 452
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 18-LIB34-086-Q1-E1-E5
US-09-960-352-4253

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 10; Length 452;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
DB 17 GCGTGGCAAGAGCGCTCTGG 37

RESULT 12
US-09-764-864-769/c
; Sequence 769, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 769
; LENGTH: 454
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-769

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 10; Length 454;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
DB 423 GCGTGGCAAGACTGCTGG 403

RESULT 13
US-09-960-352-5830
; Sequence 5830, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5830
; LENGTH: 454
; TYPE: DNA
```

```
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (181), (244), (385)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 25-LIB3058-057-Q1-K1-G1
US-09-960-352-5830

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 10; Length 454;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
DB 96 GCGTGGCAAGAGCGCTCTGG 116

RESULT 14
US-09-764-864-354/c
; Sequence 354, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 354
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-864-354

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 10; Length 498;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
DB 423 GCGTGGCAAGACTGCTGG 403

RESULT 15
US-09-962-832-224
; Sequence 224, Application US/09962832
; Patent No. US20020110821A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign
; FILE REFERENCE: 689290-74
; CURRENT APPLICATION NUMBER: US/09/962,832
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US/60/235,077
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,280
; PRIOR FILING DATE: 2000-09-25
; NUMBER OF SEQ ID NOS: 259
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 224
; LENGTH: 1443
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-962-832-224

Query Match
Best Local Similarity 77.1%; Score 16.2; DB 10; Length 1443;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTGCTGG 21
```

Tue Mar 25 09:37:26 2003

us-09-873-546-6.rnpb

Page 5

Db 425 |||||
GCCTGGGCAAGAGCCCTGG 445

Search completed: March 23, 2003, 17:17:27
Job time : 22.5425 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 14:20:51 ; Search time 136.531 Seconds
(without alignments)
2491.051 Million cell updates/sec

File: US-09-873-546-6

Perfect score: 21

Sequence: 1 gcgtggcgaagactcgtcg 21

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pla:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19.4	92.4	448	10	AM656952 109489 MA
2	19.4	92.4	454	9	AA723009 zh0e05.s
3	19.4	92.4	470	9	AI497811 tm89f05.x
4	19.4	92.4	662	10	BB648303 BB648303
5	19.4	92.4	664	12	BG699685 602681538
6	19.4	92.4	838	14	BQ887987 AGENCOURT

7	19	90.5	659	12	BE696122	BE696122 MRL-CT025
8	17.8	84.8	349	10	BB871148	BB871148 BB871148
9	17.8	84.8	368	10	AM437426	AM437426 76464 MAR
10	17.8	84.8	378	12	BF871691	BF871691 CM4-ET009
11	17.8	84.8	451	13	BI614393	BI614393 RH43755.5
12	17.8	84.8	494	13	BI362167	BI362167 RH46661.5
13	17.8	84.8	588	10	AV692927	AV692927 AV692927
14	17.8	84.8	590	10	AV689088	AV689088 AV689088
15	17.8	84.8	597	10	AV689087	AV689087 AV689087
16	17.8	84.8	609	13	BI607592	BI607592 RH74685.5
17	17.8	84.8	710	13	BI331932	BI331932 602984323
18	17.8	84.8	721	10	AV698268	AV698268 AV698268
19	17.8	84.8	729	13	BI247531	BI247531 602960169
20	17.8	84.8	770	13	BF257583	BF257583 HVMF001
21	17.8	84.8	793	13	BI144949	BI144949 602909601
22	17.8	84.8	796	12	BF384010	BF384010 602045538
23	17.8	84.8	798	13	BM048293	BM048293 603625501
24	17.8	84.8	802	12	BG775066	BG775066 602650030
25	17.8	84.8	862	14	BQ951805	BQ951805 AGENCOURT
26	17.8	84.8	982	12	BG743095	BG743095 602634230
27	17.8	84.8	1087	13	BM555453	BM555453 AGENCOURT
28	17.8	84.8	3079	11	AK004933	AK004933 Mus muscu
29	17.4	82.9	360	9	AJ473491	AJ473491 AJ473491
30	17.4	82.9	568	13	BJ013208	BJ013208 BJ013208
31	17.4	82.9	605	9	AA801711	AA801711 GM03370.5
32	17.4	82.9	605	9	A1063314	A1063314 GH03053.5
33	17.4	82.9	632	9	A1402550	A1402550 GH21792.5
34	17.4	82.9	654	13	BJ493532	BJ493532 BJ493532
35	17.4	82.9	678	13	BI240737	BI240737 RE37651.5
36	17.4	82.9	1144	17	CNS060R5	AL411843 t7 end of
37	17.4	81.0	449	17	AO846554	AO846554 LMAJFV1.1
38	16.8	80.0	139	13	BM293161	BM293161 EST575703
39	16.8	80.0	253	12	BF524549	BF524549 UT-R-AEO-
40	16.8	80.0	277	9	AA023366	AA023366 mh71g05.r
41	16.8	80.0	300	10	BE654600	BE654600 UT-M-AJ1-
42	16.8	80.0	349	6	BM375926	BM375926 EBem06-SQ
43	16.8	80.0	353	17	AQ073553	AQ073553 EP(2)2397
44	16.8	80.0	360	14	R33532	R33532 YH79C04.r1
45	16.8	80.0	365	10	BE127748	BE127748 DEPA1497

ALIGNMENTS

RESULT 1
LOCUS AM656952 448 bp mRNA linear EST 25-APR-2001
DEFINITION 109489 MARC IBOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AM656952
VERSION AM656952.1 GI:7422778
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahnensterg,S.C., Bennett,
'G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Petter,G., Holt,I., Karaymcheva,S., Liang,F., Quackenbush,J. and
Keefe,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
CONTACT: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt.-trimmed with phred

Query Match 92.4%; Score 19.4; DB 12; Length 664;
 Best Local Similarity 95.2%; Pred. No. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTCGCTGG 21
 |||||||||||
 Db 187 GCGTGGCAAGACTCGCTGG 207

RESULT 6
 LOCUS BQ887987 838 bp mRNA linear EST 16-AUG-2002
 DEFINITION AGENCOURT.8742650 NIH_MGC_129 Mus musculus CDNA clone IMAGE:6390947
 5' mRNA sequence.
 ACCESSION BQ887987
 VERSION BQ887987.1 GI:22280001
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 838)
 NIH-MGC http://mgi.nhl.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bbs-remail.nih.gov
 Tissue Procurement: Susan L. Sullivan, PhD.
 cDNA Library Preparation: Resgen, Invitrogen Corp.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M13878 row: 9 column: 12
 High quality sequence stop: 398.
 Location/Qualifiers
 1..838
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_image="6390947"
 /clone_lib="NIH_MGC_129"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: olfactory epithelium; Vector:
 pCMV-SPORT6.1 ccdB; Site_1: EcoRV; Site_2: NotI. Cloned
 unidirectionally. Primer: Oligo dt. Average insert size
 2.2 Kb. Constructed by Resgen, Invitrogen Corp. Note: this
 is a NIH_MGC library."
 BASE COUNT 150 a 277 c 227 g 184 t
 ORIGIN

Query Match 92.4%; Score 19.4; DB 14; Length 838;
 Best Local Similarity 95.2%; Pred. No. 2.1e+02;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCGTGGCAAGACTCGCTGG 21
 |||||||||||
 Db 58 GCGTGGCAAGACTCGCTGG 78

RESULT 7
 LOCUS BE696122 659 bp mRNA linear EST 11-SEP-2000
 DEFINITION MRI-C70258-040700-002-F04 C70258 Homo sapiens CDNA, mRNA sequence.
 ACCESSION BE696122
 VERSION BE696122.1 GI:10083282
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 659)

AUTHORS
 Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Briones, M. R.,
 Nagai, M. A., da Silva, M. Jr., Zago, M. A., Bordin, S., Costa, F. F.,
 Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,
 Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,
 M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and
 Simpson, A. J. G.

TITLE
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

COMMENT
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-MRI-C70258-040
 700-002-f04<3=2000-07-04<4=1)
 Seq primer: puc 18 forward
 High quality sequence start: 10
 High quality sequence stop: 599.
 Location/Qualifiers
 1..659
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="C70258"
 /dev_stage="Adult"
 /note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 716 - Ludwig Institute for Cancer Research) profiles
 into the pUC 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."
 BASE COUNT 166 a 157 c 174 g 162 t
 ORIGIN

Query Match 90.5%; Score 19; DB 12; Length 659;
 Best Local Similarity 100.0%; Pred. No. 2.9e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 GTGGGCAAGACTCGCTGG 21
 |||||||||||
 Db 37 GTGGGCAAGACTCGCTGG 55

RESULT 8
 LOCUS BB871148 349 bp mRNA linear EST 27-NOV-2001
 DEFINITION BB871148 RIKEN full-length enriched, 1 month neonate cerebellum Mus
 musculus CDNA G630027D18 5', mRNA sequence.
 ACCESSION BB871148
 VERSION BB871148.1 GI:17117358
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 349)
 Akimura, T., Arakawa, T., Carninci, P., Furuno, M., Hanagaki, T.,
 Hayatsu, N., Hirakawa, K., Hirooka, T., Hirozane, T., Imotani, K., Ishii,
 Y., Ito, M., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Matsuyama, T.,
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Okazaki, Y., Orido, T.,
 Saito, R., Sakai, C., Sakai, K., Sakazume, N., Sasaki, D., Sato, K.,
 Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa,
 A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T.,
 Wataniki, A., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.,
 RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura, T., et al.
 2001)

JOURNAL
COMMENT

Unpublished (2001)
Contact: Yoshhide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gs.c.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subcloning of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)
Wagai, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES

Source

BASE COUNT
RIGIN

68 a 110 c 114 g 57 t

RIGIN

Query Match	84.8%;	Score 17.8;	DB 10;	Length 349;
Best Local Similarity	90.5%;	Pred. No. 8.1e+02;		
Matches 19; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

QY 1 GCGTGGGCAAGAACTCGCTGG 21
| | | | | | | | | |
Db 149 GGGTGGGCAAGAGCTCGCTGG 169

Db 149 GGGTGGGCAAGAGCTCGCTGG 169

[illegible]

TITLE

JOURNAL
MEDLINE
COMMENT

Casas, E., Way, J.E., White, J., Cho, J., Fahrentz, S.C., Bennett, G.L., Heaton, M.P., Iaegeid, W., Rohrer, G.A., Chitko-Mckown, C.G., Pertea, G., Holt, I., Karanycheva, S., Liang, F., Quackenbush, J. and Keefe, J.W.

Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle Genome Res. 11 (4), 626-630 (2001)

21180013

Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 7436
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred v0.960904.e. Vectors identified by cross_match with the -m1nscore 20 and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACACGATGATGACCAT
BACKWARD: GTTTCCGAGTCACGACG
Plate: 42 row: P column: 2
Seq primer: ATTTCGTCACCTAATG.

FEATURES

Source

BASE COUNT
ORIGIN

61 a 123 c 118 g

ORIGIN

Query Match	84.8%;	Score 17.8;	DB 10;	Length 368;
Best Local Similarity	90.5%;	Pred. No. 8.2e+02;		
Matches 19; Conservative	0;	Mismatches 2;	Indels 0;	Gaps

```

QY      1 GCGTGGGCAAGACTCGCTGG 21
          |||||
Db     205 GCGTGGGCAAGAGCGCGCTGG 221

```

DB 205 GCGTGGGCAAGAGCGGCTGG 22:

RESULT	10
BE871691/c	
LOCUS	378 bp mRNA linear EST 17-JAN-2001
DEFINITION	CBA-FE0096-011100-400-F10 ET0096 Homo sapiens CDNA, mRNA sequence.
ACCESSION	BG71691
VERSION	BE871691.1 GI:12261821
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE

TITLE

JOURNAL
MEDLINE
COMMENT

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4&tl2=CM4-ET0096-011100-400-f106t3-2000-11-01&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 122
High quality sequence stop: 225.

FEATURES

Source

Location/Qualifiers

1..378

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_idb="ET0096"

/dev_stage="Adult"

/note="Organ: lung; tumor: Vector: puc18; Site:1: Sma1; Site:2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT

90 a 121 c 94 g 73 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 12; Length 378;
Best Local Similarity 90.5%; Pred. No. 8.3e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTGGCAAGACTCGCTG 21

Db 250 GCCTGGCAAGACTCGCTG 230

RESULT 11

BI614393

LOCUS

BI614393 451 bp mRNA linear EST 07-SEP-2001
RH43755.5prine RH Drosophila melanogaster normalized Head pf1c-1
Drosophila melanogaster cDNA clone RH43755 5 similar to CG8500:
Fban0008500 GO:[] located on: 3R 85E10-85E10; 08/18/2001, mRNA
sequence.

ACCESSION

BI614393.1 GI:15509918

VERSION

EST.

KEYWORDS

fruit fly.

SOURCE

Drosophila melanogaster

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 451)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
, J., Chamez, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Mista, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celinker, S. and Rubin
, G.M.

BDGP/HHMI RH Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP

TITLE

JOURNAL

COMMENT

AUTHORS

REFERENCE

FEATURES

Source

1..451

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_idb="RH43755"

/clone_idb="RH Drosophila melanogaster normalized Head

Location/Qualifiers

1..451

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_idb="RH43755"

pf1c-1"

/sex="male and female"

/dev_stage="Adult"

/lab_host="DH5-alpha Tona"

/note="Organ: head; Vector: pf1c1; Site:1: Xho1; Site:2:

BamHI; Library was kindly generated by Piero Carninci at

the RIKEN. The library was normalized and excised using

Cre recombinase. Plasmid cDNA library."

BASE COUNT

122 a 132 c 115 g 82 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 13; Length 451;
Best Local Similarity 90.5%; Pred. No. 8.8e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCCTGGCAAGACTCGCTG 21

Db 270 GCCTGGCAAGACTCGCTG 290

RESULT 12

BI362167

LOCUS

BI362167 494 bp mRNA linear EST 01-AUG-2001
RE46651.5prine RE Drosophila melanogaster normalized Embryo pf1c-1
Drosophila melanogaster cDNA clone RE46651 5 similar to CG8500:
Fban0008500 located on: 3R 85E10-85E10; 05/13/2001, mRNA sequence.

ACCESSION

BI362167.1 GI:15058195

VERSION

EST.

KEYWORDS

fruit fly.

SOURCE

Drosophila melanogaster

ORGANISM

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 494)
Stapleton, M., Brokstein, P., Hong, L., Tyler, D., Berman, B., Carlson
, J., Chamez, M., Chavez, C., Dorsett, V., Farfan, D., Frise, E., George
, R., Gonzalez, M., Guarin, H., Harris, N., Li, P., Liao, G., Mista, S.,
Mungall, C.J., Nunoo, J., Pacleb, J., Paragas, V., Park, S.,
Phouanavong, S., Wan, K., Yu, C., Lewis, S.E., Celinker, S. and Rubin
, G.M.

BDGP/HHMI RE Drosophila EST Project
Unpublished (2001)
Contact: Stapleton, M.
BDGP

TITLE

JOURNAL

COMMENT

AUTHORS

REFERENCE

FEATURES

Source

1..494

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone_idb="RE46651"

/clone_idb="RE Drosophila melanogaster normalized Embryo

pf1c-1"

/sex="male and female"

/dev_stage="0-24 hours mixed stage embryonic"

/lab_host="DH5-alpha Tona"

/note="Organ: embryo; Vector: pf1c1; Site:1: Xho1; Site:2:

BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."

BASE COUNT

133 a 148 c 124 g 89 t

ORIGIN

Query Match 84.8%; Score 17.8; DB 13; Length 494;
Best Local Similarity 90.5%; Pred. No. 9.1e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCGTGGCAGAACTCGCTGG 21
|||||
Db 313 GCGTGGCAGAACTCGCTGG 333

RESULT 13
LOCUS AV692927 588 bp mRNA linear EST 16-JAN-2002
DEFINITION AV692927 GKC Homo sapiens cDNA clone GKCD03 5', mRNA sequence.
ACCESSION AV692927
VERSION AV692927.1 GI:10294790
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 588)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
CONTACT: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
location/Qualifiers
1..588
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="GKCD03"
/clone_1lb="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 95 a 198 c 185 g 110 t

ORIGIN
Query Match 84.8%; Score 17.8; DB 10; Length 588;
Best Local Similarity 90.5%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCGTGGCAGAACTCGCTGG 21
|||||
Db 82 GCGTGGCAGAACTCGCTGG 102

RESULT 14
LOCUS AV689088 590 bp mRNA linear EST 16-JAN-2002
DEFINITION AV689088 GKC Homo sapiens cDNA clone GKCD2B06 5', mRNA sequence.
ACCESSION AV689088
VERSION AV689088.1 GI:10290951
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 590)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.

TITLE
Hu,G., Gu,J., Chen,Z. and Han,Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
CONTACT: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
location/Qualifiers
1..590
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="GKCD2B06"
/clone_1lb="GKC"
/tissue_type="hepatocellular carcinoma"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 92 a 198 c 183 g 112 t 5 others

ORIGIN
Query Match 84.8%; Score 17.8; DB 10; Length 590;
Best Local Similarity 90.5%; Pred. No. 9.6e+02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GCGTGGCAGAACTCGCTGG 21
|||||
Db 82 GCGTGGCAGAACTCGCTGG 102

RESULT 15
LOCUS AV689087 597 bp mRNA linear EST 16-JAN-2002
DEFINITION AV689087 GKC Homo sapiens cDNA clone GKCD2B05 5', mRNA sequence.
ACCESSION AV689087
VERSION AV689087.1 GI:10290950
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 597)
AUTHORS Xu,X., Huang,J., Xu,Z., Qian,B., Zhu,Z., Yan,Q., Cai,T., Zhang,X.,
Xiao,H., Qu,J., Liu,F., Huang,Q., Cheng,Z., Li,N., Du,J., Hu,W.,
Shen,K., Lu,G., Fu,G., Zhong,M., Xu,S., Gu,W., Huang,W., Zhao,X.,
Hu,G., Gu,J., Chen,Z. and Han,Z.
Insight into hepatocellular carcinogenesis at transcriptome level
by comparing gene expression profiles of hepatocellular carcinoma
with those of corresponding noncancerous liver
Proc. Natl. Acad. Sci. U.S.A. 98 (26), 15089-15094 (2001)
21625106
CONTACT: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source
location/Qualifiers
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="GKCD2B05"
/clone_1lb="GKC"
/tissue_type="hepatocellular carcinoma"

```
/dev_stage="Adult"  
/lab_host="SOLR"  
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:  
XhoI"  
BASE COUNT      96 a      196 c      182 g      117 t      6 others  
ORIGIN
```

```
Query Match      84.8%; Score 17.8; DB 10; Length 597;  
Best Local Similarity 90.5%; Pred. No. 9.6e+02;  
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY      1 GCCTGGGCAAGACTGCGCTGG 21  
      || ||||| ||||| |||||  
DB      81 GCCTGGGCAAGACTGCGCTGG 101
```

```
Search completed: March 23, 2003, 17:12:35  
Job time : 140.531 secs
```

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:37 ; Search time 694.166 Seconds
(without alignments)
1936.775 Million cell updates/sec

Title: US-09-873-546-4

Perfect score: 597

Sequence: 1 atgccgcagacagatgaacga.....gcaatgcacacctcatgtga 597

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Optical number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_101002.*
1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	583.4	97.7	1249	24	AB072525 Human MDDT encodin
2	582.8	97.6	3061	24	AB072648 Human MDDT encodin
3	380.8	63.8	1087	22	AA160838 Human polynucleoti
4	380.8	63.8	1108	22	AA159052 Human polynucleoti
5	380.8	63.8	4167	24	AD37605 Human intracellula
6	303.2	50.8	702	23	AB14159 Drosophila melanog
7	222	37.2	1286	22	AA58350 Human GTP-binding
8	222	37.2	1561	22	AAH15447 Human cDNA sequenc
9	222	37.2	3403	23	AA587856 DNA encoding novel

10	220.4	36.9	7397	19	AAV60578	Human tumour suppl
11	200.8	33.6	1481	19	AAV60577	Human tumour suppl
12	175.4	29.4	1144	22	AAH90075	Human bone marrow
13	175.4	29.4	4933	22	AAK81235	Human immune/thaema
14	175.4	29.4	4934	22	AAK81234	Human immune/thaema
15	166.4	27.9	558	11	AAQ03212	RAP2 Gene encoding
16	166.4	27.9	3300	22	AA158569	Human polynucleoti
17	160.6	26.9	2951	23	AB14158	Drosophila melanog
18	153.2	25.7	1170	23	AB103763	Drosophila melanog
19	153.2	25.7	3554	23	AB103762	Drosophila melanog
20	152.6	25.6	837	21	AAZ36913	DNA encoding a hom
21	152.6	25.6	2832	24	AB192076	Human Tumour Endot
22	152.6	25.6	2973	24	AB192087	Human Tumour Endot
23	152.6	25.6	3427	24	ABK71563	Human dltbp polynu
24	144.8	24.3	3020	24	AB192134	Mouse Tumour Endot
25	139.4	23.4	2223	22	AAH41597	Human H-Ras, c-Raf
26	139.4	23.4	2223	24	AB150184	Human H-Ras, c-Raf
27	139.4	23.4	2223	24	AB150216	Human H-Ras, c-Raf
28	139.4	23.4	2295	24	AB150201	Human H-Ras, c-Raf
29	139.4	23.4	2295	24	AB150233	Human H-Ras, c-Raf
30	137.8	23.1	570	21	AAZ50192	Ras gene for TDNE
31	137.8	23.1	570	21	AAZ50192	Human Ras cDNA, H
32	137.8	23.1	2223	22	AAH41609	Human H-Ras, c-Raf
33	137.8	23.1	2223	24	AB150196	Human H-Ras, c-Raf
34	137.8	23.1	2223	24	AB150228	Human H-Ras, c-Raf
35	137.4	23.0	657	24	ABK14232	Human related RAS
36	137.4	23.0	2223	22	AAH41606	Human R-Ras, c-Raf
37	137.4	23.0	2223	24	AB150193	Human R-Ras, c-Raf
38	137.4	23.0	2223	24	AB150225	Human R-Ras, c-Raf
39	137.2	22.6	1933	23	AB151533	Drosophila melanog
40	134.8	22.6	846	21	AAZ36892	cDNA encoding an a
41	134.8	22.6	1740	21	AAZ36891	cDNA encoding an a
42	134.8	22.6	1801	21	AAZ36893	cDNA encoding an a
43	134.8	22.6	1841	20	AAZ23022	Human Kd312 polype
44	133.2	22.3	1776	21	AAH49177	cDNA encoding huma
45	127.4	21.3	561	24	AB158432	Rat oncogenic cyto

ALIGNMENTS

RESULT 1	
AB072525	
ID	AB072525 standard; cDNA; 1249 BP.
XX	
AC	AB072525;
XX	
DT	03-SEP-2002 (first entry)
XX	
DE	Human MDDT encoding cDNA SEQ ID NO 77.
XX	
KW	Human: MDDT; disease detection and treatment molecule polynucleotide;
KW	proliferative disorder; hepatitis; psoriasis; cancer; AIDS;
KW	autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;
KW	rheumatoid arthritis; transgenic; gene therapy; antitumor sclerotic;
KW	hepatotoxic; antinflammatory; antipsoriatic; cytostatic; anti-HIV;
KW	antiallergic; antineoplastic; antidiabetic; antitumor sclerotic; antitumor;
KW	neuroprotective; antineoplastic; antitumor sclerotic; antitumor; gene; ss.
XX	
OS	Homo sapiens.
XX	
PN	W0200240715-A2.
XX	
PD	23-MAY-2002.
XX	
PF	06-SEP-2001; 2001WO-US27628.
XX	
PR	06-SEP-2000; 2000US-230505P.
PR	06-SEP-2000; 2000US-230514P.
PR	06-SEP-2000; 2000US-230515P.
PR	06-SEP-2000; 2000US-230517P.
PR	06-SEP-2000; 2000US-230518P.
PR	06-SEP-2000; 2000US-230519P.

PR	06-SEP-2000;	2000US-230865P.	
PR	06-SEP-2000;	2000US-230988P.	
PR	06-SEP-2000;	2000US-230989P.	
PR	07-SEP-2000;	2000US-230951P.	
PR	07-SEP-2000;	2000US-231163P.	
PR	07-SEP-2000;	2000US-231167P.	
XX			
PA	(INCY-)	INCYTE GENOMICS INC.	
XX			
PI	Jackson S,	Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL,	
PI	Jones AL, Yu JY, Wright RJ, Glezen D, Liu TF, Yap PE, Dahl CR,		
PI	Monlaliga MG, Bradley DL, Rohatgi SD, Harris B, Roseberry AM,		
PI	Gestin EH, Peralta CH, David MH, Panzer SR, Flores V, Daffo A,		
PI	Marwaha R, Chen AJ, Chang SC, Au AP, Imman RR;		
XX			
DR	WPI: 2002-527544/56.		
DR	P-PSDB: ABP51431.		
XX			
PT	Novel human disease detection and treatment polypeptide, useful in		
PT	diagnosis, prevention or treatment of cell proliferative disorders e.g.		
PT	arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder		
PT	e.g. AIDS		
XX			
PS	Claim 1; Page 414; 618pp; English.		
XX			
CC	The invention relates to an isolated human disease detection and		
CC	treatment (MDPT) polypeptide (I) selected from a polypeptide having a		
CC	sequence selected from 254 sequences (ABP51231-ABP51484) given in the		
CC	specification, a naturally occurring polypeptide comprising a sequence		
CC	having at least 90% identity to (I) or a biologically active or		
CC	immunogenic fragment of (I). (I) is useful for screening a compound for		
CC	effectiveness as an agonist or antagonist, for screening a compound that		
CC	specifically binds (I) or modulates the activity of (I), and for		
CC	preparing a polyclonal or monoclonal antibody by hybridoma technology.		
CC	Nucleic acids (II) (AB072449-AB072700) encoding (I) are useful for		
CC	screening a compound for effectiveness in altering expression of a target		
CC	polynucleotide comprising. Oligonucleotides and antibodies are useful for		
CC	detecting MDP in a sample or for assessing toxicity of a test compound,		
CC	in a diagnostic test for a condition or a disease associated with the		
CC	expression of MDP in a biological sample, for detecting (I) in a sample,		
CC	and for purifying (I) from a sample. A composition comprising (I), an		
CC	agonist or antagonist is useful for treating a disease or condition		
CC	associated with decreased or increased expression of functional MDP.		
CC	(I) or (II) are useful for diagnosing, treating or preventing disorders		
CC	associated with aberrant expression of MDP, where the disorders are		
CC	selected from a cell proliferative disorder such as arteriosclerosis,		
CC	cirrhosis, hepatitis, psoriasis, and cancer and an		
CC	autoimmune/inflammatory disorder such as AIDS, Addison's disease,		
CC	allergy, anaemia, asthma, arteriosclerosis, gout, multiple sclerosis or		
CC	rheumatoid arthritis. (II) are useful for creating knockin humanised		
CC	animals or transgenic animals to model human diseases. In somatic or		
CC	germline gene therapy, to generate a transcript image of a tissue or cell		
CC	type, for detecting differences in the chromosomal location due to		
CC	translocation or inversion among normal, carrier or affected individuals		
CC	and as hybridisation probes for mapping naturally occurring genomic		
CC	sequences.		
XX			
SQ	Sequence 3061 BP; 606 A; 840 C; 927 G; 688 T; 0 other:		
QY	Query Match	97.6%; Score 582.8; DB 24; Length 3061;	
QY	Best Local Similarity	99.5%; Pred. No. 2.4e-103;	
QY	Matches 595; Conservative	0; Mismatches 2; Indels 1; Gaps 1;	
DB	1	ATGCCGGAACAGATTAACGCGTGGTGTTCGGGGCGGCGGTGGGCAAG 60	
DB	51	ATGCCGGAACAGATTAACGCGTGGTGTTCGGGGCGGCGGTGGGCAAG 110	
QY	61	ACCTGCGTGTCTGCGTGTGGAAGGCGACGTTCCGGAGACCTACATCCCAACATC 120	
DB	111	ACCTGCGTGTCTGCGTGTGGAAGGCGACGTTCCGGAGACCTACATCCCAACATC 170	
QY	121	GAGGACACCTACCGGCGAGGTGATCAGCTCCGACAGAGCGTGTGACGCTGCAGATCACA 180	

Db	171	GAGGACACCTTACCGGACAG6GTGATCAGTGTGGACAAAGAGG9GTGTGCACCGCTGCAGATTCACA	230
Qy	181	GACACACCGGACGACACCAAGTTCGCCGCGATCGACGCGCTGTGCATYTCICAAAGGGCCAC	240
Db	231	GACACACCGGACGACCAACAGTTTCCGGCCATCGACGCGCTGTGCATYTCICAAAGGGCCAC	290
Qy	241	GCGTTTCATCCGTGGTTCCTCCGTCACAGCAAGCAGTCCGTGGAGAGGCTGGGGCCATC	300
Db	291	GCGTTTCATCCGTGGTTCCTCCGTCACAGCAAGCAGTCCGTGGAGAGGCTGGGGCCATC	350
Qy	301	TACAAAGCTCATCGTGAGATTCAAAGGCGAGCGTGGAGAGATCCCGGTGATGCTGTGGGC	360
Db	351	TACAAAGCTCATCGTGAGATTCAAAGGCGAGCGTGGAGAGATCCCGGTGATGCTGTGGGC	410
Qy	361	AACAAAGTCGATGTAGACGACGGGAGAGTGTAAACCGCGGAGAGGCGCGATGGGCCCA	420
Db	411	AACAAAGTCGATGTAGACGACGGGAGAGTGTAAACCGCGGAGAGGCGCGATGGGCCCA	470
Qy	421	GAGTGAAGTCCGCTTTCATGAGACCTCGCGCAAGATGAACTCAACCTCAAGAG-CT	479
Db	471	GAGTGAAGTCCGCTTTCATGAGACCTCGCGCAAGATGAACTCAACGTCAAGAGATCT	530
Qy	480	CTTCAGAGAGCTGCTGACGCTGGAGACGGCGCGGAACATGAGCCTTACATCGACGCCAA	539
Db	531	CTTCAGAGAGCTGCTGACGCTGGAGACGGCGCGGAACATGAGCCTTACATCGACGCCAA	590
Qy	540	GCGCTCCGGGAAAGAGAGAGGAGACAGACGCGTCAAGGGCAATGCAACCTCATGTGA	597
Db	591	GCGCTCCGGGAAAGAGAGAGGAGACAGACGCGTCAAGGGCAATGCAACCTCATGTGA	648
RESULT 3			
AA160838			
ID	AA160838 standard; cDNA; 1087 bp.		
AC	AA160838;		
DT	22-OCT-2001 (first entry)		
DE	Human polynucleotide SEQ ID NO 4827.		
KW	Human; neotrophic; immunosuppressant; cytostatic; gene therapy; cancer peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemolactic; chemokineic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.		
OS	Homo sapientis.		
PN	WO200153312-A1.		
PD	26-JUL-2001.		
PF	26-DEC-2000; 2000WO-US34263.		
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
PA	(HYSE-) HYSEQ INC.		
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D, Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao Qa, Zhou F, Goodrich R, Drmanac RT;		
WP1	2001: 4442253/47.		
P-PSDB	AAM41682.		


```

Db 232 AGTCCCTGGTGTGAGCTTGTGAAGGACATCCGGGAGAGCTACATCCGACGGTG 291
QY 121 GAGACACCTACACCGGCGAGTGATCAGCTGCGACAAAGAGCTGTGACCGCTGAGATACA 180
Db 292 GAAGACACCTACACCGGCGAGTGATCAGCTGTGACAAAGAGCTATGTCACATTCGACATACC 351
QY 181 GACACACCGGCGAGCGACAGTCCCGGACATGACGAGCGCTGTCCATCTCCAAAGGGGAC 240
Db 352 GACACGAGGGGAGGCCACCGCTCCCGGACATGACGCGCTGTCCATCTCCAAAGGGGAC 411
QY 241 GCCTTCATCTGCTGTGTTCTCCGTCACGACGAGTCTGCTGAGAGAGCTGGGGCCATC 300
Db 412 GCCTTCATCTGCTGTGTTCTCCATACGACGAGTCTGAGAGAGCTGACGACCATC 471
QY 301 TACAGCTCATCTGTCAGATCAAGGAGCGTGTGAGAGACATCCCGTGTGCTGTGGGC 360
Db 472 TACGAAACAAATCTGCGGATCAAGGGGAGCTGTGAGAGACATCCCGTGTGCTGTGGGC 531
QY 361 AACAGTGCATGAGACGACGAG---CGGAGGTGACACGCGCGAGGCGACGCGGTGGCC 417
Db 532 AACAGTGTATGAGAGACCGCCAGCGGAGGTGACGACGAGCGGAGCGCTTGGCC 591
QY 418 CAAGAGTGAAGTGCCTTTCATGAGAGACCTCGGCAAGATGAACCTACAACTGACAGAG 477
Db 592 CGCAGATGAGAGTGTGCTTCATGAGAGACCTCAAGCTCAACATGAAGTGAAGAG 651
QY 478 CTCTTCAGAGAGCTGCTGAGAGCTGTGAGACGCGCGAAGATGAGCTCAACATGAGAGCG 537
Db 652 CTCTTCAGAGAGCTGCTGAGAGCTGTGAGAGACGCGCGAAGATGAGCTCAACATGAGAGCG 711
QY 538 AACGCTCCGCGGAGAGAGAGAGAGAGACGCGCTCAAGGCAAAATGACACCTCATGTGA 597
Db 712 AAAAAAGCAAGCAGCAGAAAGAAAGAGAGAGCTCAAGGCAAAATGCTGTGATCATGTGA 771

RESULT 5
AAD37605
ID AAD37605 standard; cDNA; 4167 BP.
AC AAD37605;
XX
XX
XX 27-AUG-2002 (first entry)
XX
XX
XX Human intracellular signalling (INTSIG-16) cDNA.
XX
XX
XX Human; intracellular signalling; INTSIG-16; gene therapy; nausea; asthma;
XX Crohn's disease; keratosis; cell proliferative disorder; Grave's disease;
XX arteriosclerosis; hepatitis; leukaemia; autoimmune disorder; meningitis;
XX acquired immunodeficiency syndrome; AIDS; allergy; dermatitis; psoriasis;
XX Hashimoto's thyroiditis; neurological disorder; developmental disorder;
XX gastrointestinal disorder; reproductive disorder; nocturnal; antiemetic;
XX Alzheimer's disease; dementia; Creutzfeldt-Jacob disease; dermatological;
XX antiinflammatory; anaemia; nephrotropic; immunosuppressive; thyromimetic;
XX cancer; neuroprotective; cerebroprotective; antibacterial; virucide;
XX gynaecological; gene; ss.
XX
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX CDS 161..760
XX FT /*tag= a
XX FT /*product= "Human INSIG-16 protein"
XX FT sig-peptide 161..259
XX FT /*tag= b
XX FT mat-peptide 260..757
XX FT /*tag= c
XX FT /*product= "Mature INTSIG-16 protein"
XX
XX
XX W0200231152-A2.
XX
XX 18-APR-2002.

```

```

PF 12-OCT-2001; 2001WO-US32090.
XX
XX 13-OCT-2000; 2000US-240871P.
PR 30-OCT-2000; 2000US-244723P.
PR 15-NOV-2000; 2000US-249402P.
PR 22-NOV-2000; 2000US-252622P.
PR 13-DEC-2000; 2000US-255622P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Baughn MR, Ding L, Elliott VS, Gandhi AR, Gietzen KJ, Griffin JA;
XX Gururajan R, Hafalia AJA, Kearney L, Khan FA, Lal P, Lee EA;
XX Lu DM, Lu Y, Nguyen DB, Arvizu C, Ramkumar J, Tang YF;
XX Phanavetlu K, Thornton M, Walla NK, Warren BA, Xu Y, Yao MG;
XX Yue H;
XX WPI; 2002-426285/45.
XX P-PDB; AAE23389.
XX
XX Novel intracellular signalling polypeptide useful for treating and
XX preventing cell proliferative, autoimmune, neurological,
XX gastrointestinal and reproductive and developmental disorders -
XX
XX Claim 91; Page 171-172; 177pp; English.
XX
XX The invention relates to an isolated intracellular signalling (INTSIG)
XX polypeptide or a biologically active or immunogenic fragment of INTSIG.
XX INTSIG is also useful for preparing a polyclonal or monoclonal antibody
XX by hybridoma technology. INTSIG gene is useful in gene therapy. INTSIG
XX is useful for treating or preventing disorders such as cell proliferative
XX disorder e.g. keratosis, arteriosclerosis, hepatitis, psoriasis,
XX leukaemia; autoimmune disorders e.g. AIDS, allergies, anaemia, asthma,
XX Crohn's disease, dermatitis, Graves' disease, Hashimoto's thyroiditis;
XX neurological disorders e.g. Alzheimer's disease, dementia, meningitis,
XX Creutzfeldt-Jacob disease; gastrointestinal disorders e.g. nausea,
XX reproductive disorders e.g. menstrual disorders, ectopic pregnancy and
XX cancer of the breast and developmental disorders. The present sequence
XX is human INTSIG-16 cDNA.
XX
XX Sequence 4167 BP; 1098 A; 911 C; 960 G; 1198 T; 0 other;
XX
XX
XX Query Match 63.8%; Score 380.8; DB 24; Length 4167;
XX Best Local Similarity 78.3%; Pred. No. 1,7e-64;
XX Matches 470; Conservative 0; Mismatches 127; Indels 3; Gaps 1;
XX
QY 1 ATGCCGAGACAGATTAACGATTACCGCTGTGCTGGGGGGGGGGGGGGGGAG 60
Db 161 ATGCTGTAGCAGATTAACGATTACCGGGGTGCGCTGTTGGGGGGGGGGGGAG 220
QY 61 AGCTGCTGAGTGTGCGCTTGTGTAAGGAGAGTTCGCGACACCTACATCCGACATC 120
Db 221 AGCTCCCTGCTGTGAGGTTGTGAAGGACATTCGCGGAGAGCTACATCCGAGGTG 280
QY 121 GAGACACCTACCGGAGTGATCAGCTGCGACAAAGAGCTGTGACGCTGACATGACA 180
Db 281 GAAGACACCTACCGGAGTGATCAGCTGTGACAAAGACATATGACATTCGACATCAC 340
QY 181 GACACACCGGAGCGACACCACTTCCGGGCAATGACGCGCTGTCCATCTCCAAAGGGGAC 240
Db 341 GACACGAGGGGAGCCACCACTTCCGGGCAATGACGCGCTGTCCATCTCCAAAGGGGAC 400
QY 241 GCCTTCATCTGCTGTGTTCTCCGTCACGAGTCTGCTGAGAGAGCTGGGGCCATC 300
Db 401 GCCTTCATCTGCTGTGTTCTCCATACGACGAGTCTGCTGAGAGAGCTGAGCCATC 460
QY 301 TACAGCTCATCTGTCAGATCAAGGAGCGTGTGAGAGACATCCCGTGTGCTGTGGGC 360
Db 461 TACGAAACAAATCTGCGGATCAAGGGGAGCTGTGAGAGACATCCCGTGTGCTGTGGGC 520
QY 361 AACAGTGCATGAGACGACG---CGGAGGTGACACGCGCGAGGCGGAGCGGTGGCC 417
Db 521 AACAGTGTGATGAGAGCGCCAGCGGAGGTGACAGACGACGAGGCGGAGCGCTTGGCC 580

```

OY	418	CAGAGTGGAACTGCGCCTTTTCATGTGAGACACTGCGCCCAAGATGAATCAACTCAACGTCAAGAG	477
Dd	581	CGCACATGTAAGTAGTGCCCTTTCATGTGAGACCTCAAGCCAAGCTCAACATACTGAAGAG	640
OY	478	CTCTTCACAGACTGCTGTGACGTGTGAGACGCGCCGGAATCATGAGCCTCAACATCGAAGGC	537
Dd	641	CTTTTCAGAGACTGCTCAACTGTGAGAAAGCCAGAACCTGTGATCTCCAGATCGAAGGC	700
OY	538	AAGCCTCCGGAGAGCAGAAGGACAGACCGCGTCAAGGGCAAAATGCAACCCCTCATGTGA	597
Dd	701	AAAAGAGCAACAGCAGCAAAGAAAGAAAGAGAGCTCAAGAGCAAGTGGATCATGTGA	760
 RESULT 6 ABLI4159			
ID	ABLI4159	standard; cDNA; 702 BP.	
XX	AC		
XX	ABL14159;		
XX	26-MAR-2002	(first entry)	
XX	Drosophila melanogaster expressed polynucleotide seq ID NO 36959.		
XX	Drosophila: developmental biology; cell signalling; insecticide;		
XX	pharmaceutical; gene; ss.		
XX	Drosophila melanogaster.		
XX	WO200171042-A2.		
PN	27-SEP-2001.		
XX	23-MAR-2001; 2001WO-US092331.		
PF	23-MAR-2000; 2000US-191637P.		
PR	11-JUL-2000; 2000US-0614150.		
XX	(PEKE) PE CORP NY.		
PA	Venter JC, Adams M, Li PWD, Myers EW;		
P1	WPI; 2001-656860/75.		
DR	P-PsDB; ABB70056.		
XX	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -		
PT	Claim 1; SEQ ID NO 36959; 21pp + Sequence Listing; English.		
XX	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABLI6176-ABL30511), expressed DNA sequences (ABB57737-ABB72072).		
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.		
CC	Sequence 702 BP; 185 A; 187 C; 220 G; 110 T; 0 other;		
XX	Query Match	50.8%; Score 303.2; DB 23; Length 702;	
XX	Best Local Similarity	72.4%; Pred. No. 1.3e-49;	
XX	Matches 424; Conservative 0; Mismatches 153; Indels 9; Gaps		
OY	4	CCGGAACAGAGTAAACATTACCCTGCTGTGCTTTCGAGGCGCGCTGGACAAGAC	63
Dd	37	CCGAGCAGAGCAACGATTACCGCGTGTGTTTTTGAGCCGCCGCTGGACAAGAC	96
OY	64	TGCGTGTGCTGCGCTTGTGTAAGGACGTTCCGCGACACCTATACCTCCACCATCGAG	123

Db	97	TCACGTGGTGGCGCCCTTCATAAAGGCGACGTCCGGGAAAGCTACATCCCAACATCGAG	156
Oy	124	GACACCTTCGCGGAGGATCAGTCTGGACAGACGCTGCGACGCTGCAGATCACAGAC	183
Db	157	GATACGTACACACAGGTCATAGTTCGACACAAAGACATCTGCAAGCTGCAAAATCACGGAC	216
Oy	184	ACCACCGGACCCACCAAGTTCGCCGGCCATGCAGCGCCTGTCCATCTCCAAAGGCCACGCC	243
Db	217	ACCACGGGATGCATCAATTCCCGGCTATGCAGAGCGTGTCCATCTCGAAGGGGCGCGCC	276
Oy	244	TTTCATCTCGGTGTTCTCCGTCACCAGACAGTCCGTGGAGAGCTGGGGCCCATCTAC	303
Db	277	TTTCACTCCGTCCTACTCGGTGTGCTCCACAGATCTGGAGAGCTGGCGACCATCTGG	336
Oy	304	AAGCTCATCGCGCATCAAGGGCGAGCG--TGAGAGACATCCCCGTGATGCTCGGGGC	360
Db	337	GCGCTTATCAAGAGGATGTAAGGGCGCCGACATTTCCACATCTACCTTAATTTGGTGGCC	396
Oy	361	AACAAGTCGATGAGAC-----GCACCGGAGAGTGAGACACGCCGAGGCGAGCGCGGTG	414
Db	397	AACAAGTGTGATGAGACCGCGCGAATTTGAGAGAGTCTCCCAAGCGGAGGTGACGGCCAG	456
Oy	415	GCCCAAGAGTGGAGTGGCTTTTCATGAGACCTCGGCCAAGATGAACTACAAAGTCAAG	474
Db	457	GCGACCACTCTGAGCATATGCTTCATGAGACGTCGGCCAGACGAAACCAATATGTGACC	516
Oy	475	GAGCTCTTCAGAGAGCTGACGCTGGAGACCGCGGAAACATGAGCTCAACATCTGAC	534
Db	517	GAGCTGTTTCAGGAGATGCTTCACATGAGAAAGAACCGCACCTCTCCCTGCAACTGGAC	576
Oy	535	GGCACAGCGCTCCGGGAAGCAGAAAGAGACAGACCGCGCTCAAGGGCA	580
Db	577	ACCAAGAGCAGAGAGAGCAGAAAGAGAGAGAGTCCAAAGAGACA	622
RESULT 7			
ID	AAFS8350	AAFS8350 standard; cDNA; 1286 BP.	
XX	AC	AAFS8350;	
DT	19-APR-2001	(first entry)	
XX	DE	Human GTP-binding associated protein #50 coding sequence.	
KW	KW	Human: guanosine triphosphate binding associated protein; GMP; GBAF;	
KW	KW	inflammation; AIDS; Addison's disease; anaemia; arteriosclerosis; asthma	
KW	KW	autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;	
KW	KW	osteoporosis; psoriasis; ss.	
OS	OS	Homo sapiens.	
XX	PM	WO200105970-A2.	
XX	PD	25-JAN-2001.	
XX	PF	19-JUL-2000; 2000MO-US19698.	
XX	PR	19-JUL-1999; 99US-0144595.	
XX	PR	23-AUG-1999; 99US-0150460.	
XX	PR	15-OCT-1999; 99US-0159849.	
PA	PA	(INCY-) INCYTE GENOMICS INC.	
XX	PI	Yue H, Tang YT, Bandman O, Hillman JL, Lal P, Au-Young J;	
PI	PI	Reddy R, Yang J, Baughn MR, Lu DM, Azimzal Y, Patterson C;	
XX	XX	WPI: 2001-091972/10.	
DR	DR	P-PSDB: AAB68550.	
XX	XX	New guanosine triphosphate-binding associated proteins (GBAP) and their	
PT	PT	encoding nucleic acids, useful for treating and/or diagnosing diseases	

PT associated with GBAP expression, such as cancer, diabetes and asthma -
 XX
 PS Claim 5; Pages 220-221; 233pp; English.

CC The present invention relates to novel human guanosine triphosphate
 CC (GTP)-binding associated proteins (GBAPs; AAB68501-AAB68566) and their
 CC coding sequences (AAF58301-AAF58366). The proteins and coding sequences
 CC of the present invention are useful for treating a variety of disorders
 CC including inflammation, AIDS, Addison's disease, anaemia,
 CC arteriosclerosis, asthma, autoimmune disorders, Grave's disease,
 CC hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and
 CC psoriasis.

CC
 XX Sequence 1286 BP; 296 A; 355 C; 337 G; 298 T; 0 other;

Query Match 37.2%; Score 222; DB 22; Length 1286;
 Best Local Similarity 62.2%; Pred. No. 5.6e-34;
 Matches 367; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

QY 11 AGAGTAACGATTACCGCTGGTGGTGGGGGGGGGGGGGCAAGAGCTGCTGG 70
 D 456 AGATCAGAGATTACCGCTGGTGGTGGGGGGGGGGGGGCAAGAGCTGCTGG 515
 QY 71 TGTGCGCTTGTGAAGGCGACGTTCCGCGACACCTACATCCCAACATCGACACT 130
 D 516 TGCACAGTGGCGAGCGGCAACTTCCGTACATGAGTACCTCCGACATGAAATACCT 575
 QY 131 ACCGGCAGGCTATCAGCTGGCGCAAGAGCGTGTGACCGCTCAGATTCACACACACCG 190
 D 576 ACTCCAGATTCTGGCTGGCGAGCCAGGCTGTGCTTCCCTCCACATCCAGCAGCAGAGA 635
 QY 191 GCAGCCACCAAGTCCCGGCGCATGACAGCGCTGCATCTCCAAAGGCGCACCGCTTCATCC 250
 D 636 GTGGGAGAGGAGACCGGCTCTGCAGGCGCAAGTTATAGCCCGGGGCGACCGCTTCGCTCC 695
 QY 251 TGTGTTTCCGTCACGACAGCAGAGCTGTGGAGAGCTGGGGCCCATCTACACAGCTCA 310
 D 696 TGTGTTCTCAGTACCAAGAGAGAACCTGTGAAGAGCTGAAGGCTTCTATGAGCTGA 755
 QY 311 TCGTGAGATCAAGG---CAGCGTGGAGAGCATCCCGGTATGCTGTGGGCAACAGT 367
 D 756 TCTGCAGATCAAGATTAAGTAACTGTCAATAGTTCCATGTGCTGTGGCAATTAAGA 815
 QY 368 GCGATGACAGCAGCGGAGGTGTGACAGCGCGCAGGCGGCTGGCCCAAGAGTGA 427
 D 816 GTGATGACAGCCACCGGAGGTGGCCCTGATGATGATGATGATGATGATGATGATGATG 875
 QY 428 AGTGGCTTTATGAGAGCTCTGGCCAAAGATGAACTACACGTCAGAGCTTCTTCAGG 487
 D 876 ATTGCGCTTCATGAGATTTTCAAGCAAGACCGATGTAATGTGAGAGAGCTTTCACA 935
 QY 488 AGTGTGAGCGCTGAGAGCGGCGGAGACATGAGCGCTCAACATGAGAGGCAAGCGCTCG 547
 D 936 TGTGCTGTAATTACAGAGAAAAACCCACCGCGCTTCAGAGAGCCCAAGAAATCC 995
 QY 548 GGAAGCAGAGAGAGCAGACCGCTCAAGGGCAATGACACCTCATGTGA 597
 D 996 AGATGCCCAACACCACTGAGAAAGCTTGAACAAGTGCATATATATGTA 1045

RESULT 8
 AAH15447
 ID AAH15447 standard; cDNA; 1561 BP.

XX AC AAH15447;
 XX DT 26-JUN-2001 (first entry)
 XX DE Human cDNA sequence SEQ ID NO:13678.
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX OS Homo sapiens.

XX
 PN EPI074617-A2.

XX
 PD 07-FEB-2001.

XX
 PF 28-JUL-2000; 2000EP-0116126.

XX
 PR 29-JUL-1999; 99JP-0248036.

XX
 PR 27-AUG-1999; 99JP-0300253.

XX
 PR 11-JAN-2000; 2000JP-0118776.

XX
 PR 02-MAY-2000; 2000JP-0183767.

XX
 PR 09-JUN-2000; 2000JP-0241899.

XX
 PA (HELI-) HELIX RES INST.

XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX
 PT WPI; 2001-318749/34.

XX
 PS Claim 8; SEQ ID 13678; 2537pp + CD ROM; English.

CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 1561 BP; 422 A; 360 C; 373 G; 406 T; 0 other;

Query Match 37.2%; Score 222; DB 22; Length 1561;
 Best Local Similarity 62.2%; Pred. No. 5.7e-34;
 Matches 367; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

QY 11 AGAGTAACGATTACCGCTGGTGGTGGGGGGGGGGGGGCAAGAGCTGCTGG 70
 D 505 AGATCAGAGATTACCGCTGGTGGTGGGGGGGGGGGGGCAAGAGCTGCTGG 564
 QY 71 TGTGCGCTTGTGAAGGCGACGTTCCGCGACACCTACATCCCAACATCGAGAGCACT 130
 D 565 TGCACAGTGGGGCGAGGCGCAACTTCGTCATGATTAAGTACCTGCGACCAATGAATAATCT 624
 QY 131 ACCGGCAGGCTATCAGCTGGCGCAAGAGCGTGTGACAGCTGCAGATTCACAGACACCG 190
 D 625 ACTGCAGATTGCTGGCTGCGAGCCAGCGTGTGCTTCCCTGACATCCAGCAGCAGAGA 684
 QY 191 GCAGCAGCAAGTCCCGGCGCATGACAGCGCTGTCCATCTCCAAAGGCGCACCGCTTCATCC 250
 D 685 GTGGGAGCGCAACCGCGCTCTGCGAGCGCAAGTTATAGAGCGGGGCCACCGCTTCGCTCC 744
 QY 251 TGTGTTCTCCGTCACGACAGCAGTGTGAGAGAGCTGGGGCCCATCTACACAGCTCA 310

```

Db      745 TGGTCTACTAGTCACCAAGAGAAACCTGGAAAGCTGAAGGCTTTATGAGCTGA 804
QY      311 TCGTGCAGCAAGG---CAGCGTGAGGACATCCCGTGATGCTCGGCAACAAGT 367
Db      805 TCTGCAAGATCAAGATCAACACCTGCATTAAGTTCCTCCATCGTGTGGCAATTA 864
QY      368 GCGATGAGACGACGCGGAGGTGACACGCGGAGCGGCGGTGGCCCAAGAGTGA 427
Db      865 GTATGATCAACCCACCGGAGAGTGGCCCTGAATGATGTGCTCCACCTGCGATGAGTGA 924
QY      428 AGTGCCTTTCATGAGACCTCGGCAAGTGAACATCAACGCTCAAGACCTCTCCAG 487
Db      925 ATTGCGCTTCATGAGATTTCAAGCCAGACCGATGTGAATGTGACGACCTGTCCGA 984
QY      488 AGCTGCTGACGTGAGACCGCGGACATGAGCCTCAACATGACGCGCAACGCTCCG 547
Db      985 TCGTGTGAATTAAGAAAGAAAGCCACCGCGCTCCGAGAGCCCGAAGAAATCC 1044
QY      548 GGAAGCAGAAGAGACAGACCGCGCTCAAGGCGCAATGCAACCTCATGTGA 597
Db      1045 AGATGCCAACACCACTGAGAGCTGCTTGAACAAGTGCAATATCATGTGA 1094

```

RESULT 9

AA587856 standard; cDNA; 3403 BP.

AA587856;

13-FEB-2002 (first entry)

DNA encoding novel human diagnostic protein #23660.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001MO-US08631.

31-MAR-2000; 2000US-0540217.

23-AUG-2000; 2000US-0649167.

(HYSE-) HYSEQ INC.

Dormanac RT, Liu C, Tang YT;

WPI: 2001-639362/73.

P-PSDB: ABG23669.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 1: SEQ ID NO 23660; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating

disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. A564197-A594564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.int/pub/published_pct_sequences.

Sequence 3403 BP; 909 A; 850 C; 887 G; 757 T; 0 other;

Query Match 37.2%; Score 222; DB 23; Length 3403;

Best Local Similarity 62.2%; Pred. No. 5, 9e-34;

Matches 367; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

```

QY      11 AGAGTAAACGTTACCGCGTGTGTGTCGGGCGGCGGCGTGGCAAGAGCTCGCTG 70
Db      2814 AGATCAGAGATTAACCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2873
QY      71 TCGTGCCTTCGTGAAGGCGACGCTCCGACACATCAATCCACATGAGAGACCT 130
Db      2874 TGCACAAAGTGGCGGAGCGCAATTCCTCATAGTACCTGCGACATTGAAATACCT 2933
QY      131 ACCGCAAGTGTATGACGTGCGACAGACGCTGTGACGCTGCAATCAGACACCAACC 190
Db      2934 ACTGCCAGTTGTGCGCTGCGACGACGCTGTGCTTCCCTGCACATACCGACGACAGA 2993
QY      191 GCAGCCACCACTTCCCGGCGATGACGCGCTGTCATCTCCATCCAGCCGCGCTTATCC 250
Db      2994 GTGGCGAGCGCAACCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3053
QY      251 TGGTGTTCCTCGTACACGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 310
Db      3054 TGGTCTACTAGTCACCAAGAGAAAGAAACCTGGAAGAGCGCTTATGAGACTGA 3113
QY      311 TCGTGCAGATCAAGG---CAGCGTGAGGAGACATCCCGTGATGCTGTGTGGCAACA 367
Db      3114 TCTGCAAGATCAAGGATTAACACCTCATTAAGTTCCTCCATCGCTGTGTGGCAAT 3173
QY      368 GCGATGAGACGCGGAGGAGTGACACGCGGCGGAGGCGGAGGCGGCGGCGGCGGCGG 427
Db      3174 GTATGACACCCACCGGAGAGTGGCTGTGAATGATGTGTCACCTGTGCGATGAGTGA 3233
QY      428 AGTGCCTTTCATGAGACCTCGGCCAAGATGAACATAACGTCAGAGAGCTTTCAG 487
Db      3234 ATTGCGCTTCATGAGATTTTCAAGCAAGACCGATGATGTGAGAGCTTTCAC 3293
QY      488 AGCTGCTGACGCTGAGACGCGCGGACACATGAGCCTCAACATGACGCGCAAGCGCT 547
Db      3294 TCGTGTGAATTAAGAAAGAAAGCCACCAACCGCTCCAGAGGCGGAGAAATATCC 3353
QY      548 GGAAGCAGAAGAGACAGACCGGCTCAAGGCGCAATGCAACCTCATGTGA 597
Db      3354 AGATGCCAACACCACTGAGAGAGCTGCTTGAACAAGTGCATATCATGTGA 3403

```

RESULT 10

AAV60578 standard; cDNA; 7397 BP.

AAV60578;

18-JAN-1999 (first entry)

Human tumour suppressor gene NOEY2.

NOEY2; tumour suppressor; human; breast cancer; ovary cancer; diagnosis; gene therapy; transgenic animal; ss.

Homo sapiens.

FH Key Location/Qualifiers
 FT CDS 5468..6158
 FT /*tag= a
 XX
 PN MO9842830-A2.
 PD 01-OCT-1998.
 XX
 PD 20-MAR-1998; 98WO-US05723.
 PF
 XX 13-JAN-1998; 98US-0071263.
 PR 21-MAR-1997; 97US-0041580.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Bast RC, Xu F, Yu Y;
 XX WPI; 1998-532005/45.
 DR P-PSDB; AAW/9898.
 XX
 XX New nucleic acid encoding NOEY2 tumour suppressor from ovarian
 PT epithelium - useful for, e.g. treatment, diagnosis and prognosis of
 cancer, particularly cancer of ovary and breast
 PS Claim 4; Page 152-158; 182pp; English.
 XX
 XX This is the NOEY2 gene nucleotide sequence. NOEY2 is a novel human
 CC tumour suppressor gene that is expressed in normal ovarian and
 CC breast surface epithelial cells but is consistently absent or
 CC down-regulated in ovarian and breast cancer cells. The open
 CC reading frame encodes a 26 kDa polypeptide (see AAW/9898) that shows
 CC homology to members of the Ras and Rap family. The gene is
 CC isolated from human chromosome 1p31. A full-length cDNA clone
 CC (see AAV60577) is also claimed. NOEY2 nucleic acids can be used in
 CC the recombinant production of NOEY2 polypeptides, in the gene
 CC therapy of cancer (especially ovarian and breast-related), and to
 CC prepare transgenic animals useful as models of cancer. Fragments
 CC of NOEY2 nucleic acids are also used as probes and primers for
 CC diagnosis, including detection of mutations, or as antisense
 CC molecules or ribozymes for reducing/eliminating NOEY2 activity. The
 CC nucleic acids can also be used to screen for antitumour agents that
 CC stimulate NOEY2, overcome lack of this protein or block expression
 CC of mutant NOEY2.
 XX
 SQ Sequence 7397 BP; 2043 A; 1542 C; 1651 G; 2063 T; 98 other:
 Query Match 36.9%; Score 220.4; DB 19; Length 7397;
 Best Local Similarity 62.0%; Pred. No. 1.3e-33;
 Matches 366; Conservative 0; Mismatches 221; Indels 3; Gaps 1;
 11 AGAGTAACGATTAACCGGCTGGTGGTGGGGGGGGGGGCAAGAGCTCGCTGG 70
 11
 Db 5568 AGATCAGAGATTACCGGCTGTGTAGTCCGACCCGCTGGTGGGAAAGTACGGCTGC 5627
 71 TGTGCGCTTCGTGAGGGGCACTTCCGCGACACTTACATCCCAACATCAGAGACACT 130
 11
 Db 5628 TGCACAGATGGGGGCAACTCCGTATGATAGTACCTCCACCATTAATAATACCT 5687
 131 ACCGGCAGGTGATCAGCTCGCAGCAAGAGCGTGTGCAGCGCTGCAGATCAGACACCACCG 190
 11
 Db 5688 ACTGCCAGTGTGCTGGCTGAGCCAGCGTGTGCTTCCCTGCATACACCAAGCAGCAAGA 5747
 191 GCAGCCACAGTTCCCGCCATGACAGCGCTGTCCATCTCCAAAGGCCACGCCCTTCATCC 250
 11
 Db 5748 GTGGCGAGCGCAACCGCTGTGAGCGCCAGTTATAGCCCGGGGCGACGCTTCGCTCC 5807
 251 TGTGCTTCCTCCATCAGCAGCAAGCAGTCCGAGAGAGCTGGGCGCCATCAACAGTCA 310
 11
 Db 5808 TGTGCTCTAGTACACCAAGAGAAACCCGTGAAAGACCTTAAGCCCTTCTATAGCTGA 5867
 311 TGTGTCAGATCAAGG---CAGCGTGAAGAGCATCCCGGTGATCTCGTGGGCAACAAGT 367
 11
 Db 5868 TGTGCAAGATCAAGATCAACAACTGCATAGTTCCTCCATGCTGCTGGTGGCAATAAA 5927

368 GCGATGAGACGCGAGCGGAGGNGACACGCGCGAGGCGCGGCGCCCAAGAGTGA 427
 11
 Db 5928 GTGATGACACCCACCGGAGGTGGCCCTGATGATGTCGCCACCTGTCGATGGATGGA 5987
 428 AGTGGCTTCCTGATGAGAGCCTCGGCGCAAGATGAACCTACAGCTCAAGAGAGCTTCCAGG 487
 11
 Db 5988 ATTGGGCTTCATGAGATTTACAGTAAAGACCGATGATGATGATGAGAGAGTGTTCACA 6047
 488 AGCTGCTACCGCTGGAGACGCGCGGAGAACATGAGCCTCAACATGAGCGGCAAGCGCTCG 547
 11
 Db 6048 TGTGCTGAATTAACAGAAAAAGCCACCGGCTCCAGAGCCCAAGAAATCC 6107
 548 GGAAGCAGAGAGAGACAGCCGTCAGAGGCAATGCAACCTCATGTGA 597
 11
 Db 6108 AGATGCCACACCACTGAGAAAGCTTGGACAAAGTGCATATCATGTGA 6157
 RESULT 11
 AAV60577
 ID AAV60577 standard; cDNA; 1481 BP.
 XX
 AC AAV60577;
 XX
 DT 18-JAN-1999 (first entry)
 XX
 DE Human tumour suppressor gene NOEY2 cDNA.
 XX
 XX NOEY2; tumour suppressor; human; breast cancer; ovary cancer;
 KW diagnosis; gene therapy; transgenic animal; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 150..836
 FT /*tag= a
 XX
 XX MO9842830-A2.
 FT 01-OCT-1998.
 XX
 PF 20-MAR-1998; 98WO-US05723.
 XX
 PR 13-JAN-1998; 98US-0071263.
 PR 21-MAR-1997; 97US-0041580.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Bast RC, Xu F, Yu Y;
 XX WPI; 1998-532005/45.
 DR P-PSDB; AAW/9898.
 XX
 XX New nucleic acid encoding NOEY2 tumour suppressor from ovarian
 PT epithelium - useful for, e.g. treatment, diagnosis and prognosis of
 cancer, particularly cancer of ovary and breast
 PS Claim 4; Page 148-150; 182pp; English.
 XX
 XX This is the complete NOEY2 cDNA nucleotide sequence. NOEY2 is a
 CC novel human tumour suppressor gene that is expressed in normal
 CC ovarian and breast surface epithelial cells but is consistently
 CC absent or down-regulated in ovarian and breast cancer cells.
 CC The open reading frame encodes a 26 kDa polypeptide (see AAW/9898)
 CC that shows homology to members of the Ras and Rap family.
 CC Differential display PCR was used to isolate and clone the novel
 CC gene, and RACE was used to amplify the full-length NOEY2 cDNA from
 CC human ovarian surface epithelial cells. A NOEY2 genomic DNA
 CC sequence (see AAV60578), isolated from human chromosome 1p31, is also
 CC claimed. NOEY2 nucleic acids can be used in the recombinant
 CC production of NOEY2 polypeptides, in the gene therapy of cancer
 CC (especially ovarian and breast-related cancers), and to prepare
 CC transgenic animals useful as models of cancer. Fragments of NOEY2

CC nucleic acids are also used as probes and primers for diagnosis,
 CC including detection of mutations, or as antisense molecules,
 CC ribozymes for reducing/eliminating NOB2 activity. The nucleic
 CC acids can also be used to screen for antitumour agents that
 CC stimulate NOB2, overcome lack of this protein or block expression
 CC of mutant NOB2.

SO Sequence 1481 BP; 405 A; 329 C; 360 G; 387 T; 0 other;

Query Match 33.6%; Score 200.8; DB 19; Length 1481;
 Best Local Similarity 62.2%; Pred. No. 6.8e-30;
 Matches 367; Conservative 0; Mismatches 217; Indels 6; Gaps 3;

QY 11 AGAGTACATTAACCGGCTGGTGTGTTGGGGGCGGCGGCGGCAAGACTCGCTG 70
 DB 250 AGATTCAGATTAACCGGCTGGTGTGTTGGGGGCGGCGGCGGCAAGACTCGCTG 309
 QY 71 TCGTGGCTTCGTGAAGGCGACGTTCCGCGACACCTACATCCCGACATGAGACACCT 130
 DB 310 TGCACAGTGGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 369
 QY 131 ACCGCGAGTGTATGACCTGCGACAGAGCGGTGACGCGTGCAGATACAGACCGACCG 190
 DB 370 ACTGCAGATTGCT-GGCTGACGCGCGGTGTGCTTCCCTGCACATACCGACAGCAAGA 428
 QY 191 GCAGCCACAGTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 250
 DB 429 TGGCGGCGAC--GGCAGCGGCTGCGACGCGGCGGCGGCGGCGGCGGCGGCGG 486
 QY 251 TGGTGTTCCTCCGACACAGACATCGCTGGAGAGCGGCGGCGGCGGCGGCGGCGG 310
 DB 487 TGGTGTTCCTCCGACACAGACATCGCTGGAGAGCGGCGGCGGCGGCGGCGGCGG 546
 QY 311 TCGTGCAGATCAAGG--CAAGCGTGAAGACATCCCGGTGATGCTGGGCGCAACAGT 367
 DB 547 TCGCAAGATCAAGG--CAAGCGTGAAGACATCCCGGTGATGCTGGGCGCAACAGT 606
 QY 368 GCGATGAGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 427
 DB 607 GTCGATACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 666
 QY 428 AGTGCCTTTCAATGAGACCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 487
 DB 667 ATTGCCTTTCAATGAGACCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 726
 QY 488 AGCTGTGACGCTGGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 547
 DB 727 TCTGTGATTAACAGAAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 786
 QY 548 GGAAGCAGAAAGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 597
 DB 787 AGATGCCCAACACCACTGAGAGCTGCTGACAGATGATTAATCATGTGA 836

RESULT 12
 AAH90075
 ID AAH90075 standard; cDNA; 1144 BP.
 XX AAH90075;
 DT 01-OCT-2001 (first entry)
 XX Human bone marrow cDNA, SEQ ID NO: 319.
 DE
 XX Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
 KM antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
 KM immunosuppressive; gene therapy; cytokine cell proliferation;
 KM cell differentiation modulator; immune disorder; infection; cancer;
 XX human immunodeficiency virus; HIV; autoimmune disorder; haemophilia; ss.
 OS Homo sapiens.
 XX
 XX
 PM WO200153453-A2.

XX 26-JUL-2001.
 PD 23-DEC-2000; 2000MO-US34960.
 XX
 XX
 PF 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-052317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 30-NOV-2000; 2000US-0250583.

XX (HYSE-) HYSEQ INC.

PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
 PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Dirmannac RT;

DR WPI: 2001-488707/53.
 DR P-PSDB: AAM00956.

PT Novel bone-marrow-expressed polynucleotides and polypeptides, useful
 for treating e.g. cancer and immune deficiency disorders -
 Claim 1; Page 425; 648pp; English.

The present sequence is one of 251 novel human polynucleotides
 expressed in the bone marrow. The polynucleotide and the
 polypeptide encoded by it are useful in the treatment of various
 CC immune deficiencies and disorders. The deficiencies and disorders may
 CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
 CC infection, or may result from an autoimmune disorder, a coagulation
 CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
 CC suppression of an inflammatory response or treatment of a nervous
 CC system disorder such as Alzheimer's disease. Detection of the presence
 CC or increased expression of the polynucleotide or the protein it
 CC encodes is useful for the diagnosis and/or prognosis of one
 CC or more types of cancer. The polynucleotide and polypeptide can be
 CC used as nutritional sources or supplements and in the screening of
 CC chemical compounds as potential drugs.

SO Sequence 1144 BP; 220 A; 314 C; 367 G; 243 T; 0 other;

Query Match 29.4%; Score 175.4; DB 22; Length 1144;
 Best Local Similarity 58.2%; Pred. No. 5.1e-25;
 Matches 308; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 19 GATTACCGGCTGGTGTGTTCCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 78
 DB 109 GAGTACAAAGGTGGTGTGCTGCTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 168
 QY 79 TTGTAAGGCGACGCTCCGCGACACCTACATCCCGGCGGCGGCGGCGGCGGCGG 138
 DB 169 TTGTAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 228
 QY 139 GTGATCAGCTGCAGAAAGCGGTGTCACAGCTGACAGATACAGACCAACCGGACCC 198
 DB 229 GAGATTGAGGTGAGCTGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 288
 QY 199 CAGTTCGCGCTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 258
 DB 289 CAGTTCGCGCTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 348
 QY 259 TCCGTCACAGCAAGCAAGTCCGTCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 318
 DB 349 ACCCTGCTAACAGCAGAGCTTCCAGGACATCAAGCCATCCGGGACCAAGTATCCGC 408
 QY 319 ATCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 378
 DB 409 GTGAAGCGGTACAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 468

PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -

Disclosure; SEQ ID NO 36046; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to
supplement the patient's own production of (I). Additionally, (I)
polynucleotides may be used to produce the secreted (I), by inserting
the nucleic acids into a host cell and culturing the cell to express the
protein. (I) proteins and polynucleotides may be used to prevent,
diagnose and treat immune/hematopoietic-related diseases, especially
cancers and cancer metastases of hematopoietic-derived cells. AAK64703
to AAK87694 represent human immune/hematopoietic antigen genomic
sequences from the present invention. AAK54942 to AAK54950 and AAM82169
represent sequences used in the exemplification of the present invention.

Sequence 4934 BP; 1542 A; 1175 C; 985 G; 1232 T; 0 other;

Query Match 29.4%; Score 175.4; DB 22; Length 4934;
Best Local Similarity 58.2%; Pred. No. 5; Se-25;
Matches 308; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

OY	19	GATTTCACCGGTGGTGGTTCTTGCGGAGCCGCCTGGGCAAGAAGCTCCTGGTGGTCCGCC	78
Db	4470	GAGTACAAAGTGTGTGTCTCTGGCTCGGGGCGGTGGGCAATGCCGCTCACCCGTGACG	4411
OY	79	TTCGTGAAGGACGCTTCCGGACACCCTACATCCCCACATCGAGACACTACCGGCAG	138
Db	4410	TTTCGTAGAGGGGTCTTCATCATGAAAGTACGACCCGACATCGAACAATCTTTAACGCCAAG	4351
OY	139	GTGATTCAGCTCGGACAGAGCGTGTGCACCGCTGCATGATCACAAGACACCAACGCGCAC	198
Db	4350	GAGATTGAGGTGGACTCTCTCGCCCTCGTGTGTGAGATCTCTGTGATACGGCGGGCACCGAG	4291
OY	199	CAGTTCCCGCCGACACGACCGCTCTCATCTCCAAAGGCCACGACCTTCAATCCTGTGTTTC	258
Db	4290	CAGTTCCGGCTCTCATGCGGGACCTTACATCAAGAACGGCCAGGCTTCAATCCTGTGTTAC	4231
OY	259	TCCGTTCACCAACAGCACTGCTGTGAGAGCTGGGGCCCATCTACAACTCATCTGTGACG	318
Db	4230	AACCTCTGTCAACCAAGCAGAGCTTCCAAGACATCAAGCCATGTGGGACCAAGATCATCTCCG	4171
OY	319	ATCAAGGGCAGCGTGTGAGAGCATCTCCCCTGATCTCTGTGGGCAACAAGTGCATGAGACG	378
Db	4170	GTGAAGCGGTACGAGCGCGTGCCTCATGATCTCTGTGGGCAACAAGTGSACCTGGAGGAT	4111
OY	379	CAGCGGAGGTGAGACACGCGCGGAGGCGGAGGCGGTGGCCCAAGAGTGAAGTCCCTTC	438
Db	4110	GAGCGCCAGAGTCTGTGTACGGGAGGACCAAGGCTCTGAGAGAGTGGAGCTGCCCTTC	4051
OY	439	ATGGAACCTCTGGCCCAAGATGAACATACACGTCACAGAGACTCTCCAGAGAGCTCTGACG	498
Db	4050	ATGGAACGCTGTGGCCCAAAACAAAGCCTCGGTGACAGACTTAATTTGCCAAGATCGTGGG	3991
OY	499	CTGGAGACGCCCGGAAACATGAGCTTAACATGACGCGCAAGCGCTCGG	547
Db	3990	CAGATGAACCTACGCGCGGACCAAGCCCAACGCGATGAGAGCTCGTCTCGG	3942
RESULT 15			
ID	AAO03212	standard; DNA; 558 BP.	
XX	AAO03212;		
DT	10-JUL-1990	(first entry)	
DE	RAP2 Gene encoding at least one peptide associated with ras oncogene.		
KW	ras oncogene; cancer; GTPase; ds.		
XX	Synthetic.		
FH	Key	Location/Qualifiers	
FT	CDS	4..552	
XX		/*lag= a	
PV	WO9000179-A.		
PD	11-JAN-1990.		
PF	04-JUL-1989;	89WO-FR00348.	
PR	04-JUL-1988;	88FR-0009031.	
PA	(INRM) INSERM INST NAT SANTE.		
PI	Tavitlan A, Pilzon V, Chardin P;		
DR	WPI; 1990-037122/05.		
DR	P-PsDB; AAR05076.		
PT	DNA encoding sequences associated with human Oncogenes -		
TF	and derived antibodies, useful for in vitro diagnosis of cancer		
TF	and for therapy.		

[illegible]

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 14:20:51 ; Search time 3881.38 seconds
(without alignments)
2491.051 Million cell updates/sec

Title: US-09-873-546-4

Sequence: 1 atcgccgacagcagatgaaga.....gcaatgcacccctatgtga 597

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

otal number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estinu:*
5: em_estrov:*
6: em_estipl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estrom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	525.4	88.0	664	12	BG699685 602681538
2	470	78.7	662	10	BB648303 BB648303
3	454.6	76.1	838	14	BO887987 AGENCOURT
4	416.4	69.7	470	9	A1497811 tm89f05.x
5	402.4	67.4	454	9	AA723009
6	386.6	64.8	458	10	BE095820 UI-R-BUO-

c	7	381.4	63.9	456	10	BE095936	BE095936 UI-R-BUO-
c	8	353.8	59.3	422	9	A1764459	A1764459 UI-R-YO-a
c	9	349.2	58.5	419	10	AM533412	AM533412 UI-R-BUO-
c	10	339.4	56.9	419	10	BG702913	BG702913 602684858
	11	336.6	56.4	645	13	B1547693	B1547693 603191841
	12	328.4	55.0	702	10	AM305720	AM305720 f63a05.y
	13	322.2	54.0	786	13	B1552244	B1552244 603195901
	14	312.2	52.3	628	13	BM172422	BM172422 imaggc.4
	15	311.2	52.1	709	12	BG699614	BG699614 602679344
	16	302.2	50.6	632	13	B1475251	B1475251 f930a09.y
	17	302	50.6	626	10	AM134357	AM134357 f118011.y
	18	301.6	50.5	618	9	AU171413	AU171413 AU171413
	19	294.2	49.3	893	14	BO922385	BO922385 AGENCOURT
	20	281.8	47.2	466	10	BE198488	BE198488 bb60e12.y
	21	265.6	44.5	462	12	BF193212	BF193212 244717.MA
	22	261.8	43.9	573	13	BM186599	BM186599 fV73a03.y
	23	246.8	41.3	558	13	BM034832	BM034832 fu36f03.y
	24	237.8	39.8	478	10	AM784881	AM784881 115197.MA
	25	237.6	39.8	448	10	AM656952	AM656952 109489.MA
	26	228.2	38.2	687	14	BM944010	BM944010 UI-M-EMDP
	27	220.4	36.9	467	12	BC9892366	BC9892366 fd75d04.y
	28	213.2	35.7	349	10	BB871148	BB871148 BB871148
	29	209.2	35.0	1141	13	BM544956	BM544956 AGENCOURT
	30	204.2	34.2	756	12	BG707443	BG707443 602672864
	31	203.8	34.1	847	9	AL668307	AL668307 AL668307
	32	203.4	34.1	1119	12	BE885242	BE885242 601506309
	33	203.2	34.0	908	13	BI603847	BI603847 603249732
c	34	202.8	34.0	961	9	AL664187	AL664187 AL664187
	35	200.4	33.6	550	13	BM263033	BM263033 1925611.y
	36	195.4	32.7	840	12	BF213048	BF213048 601814279
	37	192.4	32.2	560	13	BI539761	BI539761 452552.MA
	38	192.4	32.2	565	13	BM107153	BM107153 511299.MA
	39	189.4	31.7	681	9	AL667679	AL667679 AL667679
	40	189.2	31.7	810	13	BI599064	BI599064 603250464
	41	188.6	31.6	895	9	AL666801	AL666801 AL666801
	42	187	31.3	727	12	BG776910	BG776910 602664101
	43	185.2	31.0	799	13	BI668300	BI668300 603295661
	44	178.2	29.8	497	13	BI849738	BI849738 477665.MA
c	45	178.2	29.8	591	10	BE605570	BE605570 f113a08.x

ALIGNMENTS

RESULT 1
LOCUS BG699685 664 bp mRNA linear EST 07-MAY-2001
DEFINITION 602681538F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4814337 5',
mRNA sequence.
ACCESSION BG699685
VERSION BG699685.1 GI:13968244
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 664)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-f@mail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroaki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10709 row: k column: 10
High quality sequence stop: 662.
Location/Qualifiers

source

1. .664

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="4814337"

/clone_lib="NH_MGC_95"

/tissue_type="hippocampus"

/lab_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag) ; Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTTT-3', size-selected for average insert size 2.5 kb and normalized to ROT 5. This is a primary library enriched for full-length clones and constructed using the cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: this is a NIH-MGC Library."

BASE COUNT 131 a 212 c 229 g 92 t

ORIGIN

Query Match 88.0%; Score 525.4; DB 12; Length 664;

Best Local Similarity 99.8%; Pred. No. 7,6e-96;

Matches 526; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATCCGGAGACAGATTAACGATTACCGCGTGTGTGTTGGGGGGGGCGGCGGCGCAG 60

138 ATCCGGAGACAGATTAACGATTACCGCGTGTGTGTTGGGGGGGGCGGCGGCGCAG 197

61 ACCTGCGT 120

198 ACCTGCGT 257

121 GAGGACACCTACCGGAGGTGATGCTGCGACAGAGGCGTGTGTGTGTGTGTGTGTGT 180

256 GAGGACACCTACCGGAGGTGATGCTGCGACAGAGGCGTGTGTGTGTGTGTGTGTGT 317

181 GACACACCTACCGGAGGTGATGCTGCGACAGAGGCGTGTGTGTGTGTGTGTGTGT 240

318 GACACACCTACCGGAGGTGATGCTGCGACAGAGGCGTGTGTGTGTGTGTGTGTGT 377

241 GCCTTCATCCTGCTGT 300

378 GCCTTCATCCTGCTGT 437

301 TACAAGCTCATGCTGATGATCAAGGCGAGCGTGTGTGTGTGTGTGTGTGTGTGTGT 360

438 TACAAGCTCATGCTGATGATCAAGGCGAGCGTGTGTGTGTGTGTGTGTGTGTGTGT 497

361 AACAAAGTCGATGAGACGACGCGGAGGTGACACGCGGAGCGCGAGCGGTGCGCCAA 420

498 AACAAAGTCGATGAGACGACGCGGAGGTGACACGCGGAGCGCGAGCGGTGCGCCAA 557

421 GAGTGAAGTGGGCTTTTCATGAGACCTTCGCGCAAGATGAACTACACGCTCAAGAGCTC 480

558 GAGTGAAGTGGGCTTTTCATGAGACCTTCGCGCAAGATGAACTACACGCTCAAGAGCTC 617

481 TTCCAGAGCTGCTGAGCGTGGAGCGCGCGGAGCATGAGCGCTCAA 527

618 TTCCAGAGCTGCTGAGCGTGGAGCGCGCGGAGCATGAGCGCTCAA 664

RESULT 2

BB648303 662 bp mRNA linear EST 26-OCT-2001

BB648303 LOCUS BB648303 RIKEN full-length enriched, 16 days embryo head Mus

DEFINITION BB648303 RIKEN full-length enriched, 16 days embryo head Mus

ACCESSION BB648303.1 GI:16482559

VERSION BB648303.1

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 662)

AUTHORS TITLE JOURNAL COMMENT

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Saito, R., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suenho-cho, Tsukuba, Ibaraki, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp

URL: http://genome.gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

wagl, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer. Genome Res. 10 (11), 1757-1771 (2000)

Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamana, I., Alizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and Hayashizaki, Y.

Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES source

1. .662

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone_image="C130007J02"

/clone_lib="RIKEN full-length enriched, 16 days embryo head"

/sex="mixed"

/tissue_type="head"

/dev_stage="16 days embryo"

/lab_host="DH10B"

/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATTCGACGATTAATTAATATCCGCCGCCGCC 3']. cDNA was prepared by using thermostable thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGACGATTAATTAATTAATATCCGCCGCCGCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I"

BASE COUNT 144 a 191 c 222 g 105 t

ORIGIN

SOURCE	ORGANISM
human.	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.


```

OY 301 TACAAGCTCATGTCAGATCATCAAGGCGAGCTGGAGGACATCCCGTATGCTGTGGC 360
      |||||||
Db 349 TACAAGCTCATGTCAGATCATCAAGGCGAGCTGGAGGACATCCCGTATGCTGTGGC 408
OY 361 AACAGTCCGATGAGACGCGGAGGTGAGACACGCCGAGGC 404
      |||||||
Db 409 AACAGTCCGATGAGACGCGGAGGTGAGACACGCCGAGGC 452

RESULT 6
BE095820/c 458 bp mRNA linear EST 12-JUN-2000
LOCUS      UI-R-BU0-a02-h-08-0-UI.s1 UI-R-BU0 Rattus norvegicus cDNA clone
DEFINITION UI-R-BU0-a02-h-08-0-UI 3', mRNA sequence.
ACCESSION  BE095820
VERSION     BE095820.1 GI:8486751
KEYWORDS    EST.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 458)
            Bonaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: mesoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA-No.
FEATURES
    source          1..458
                    /organism="Rattus norvegicus"
                    /strain="Sprague-Dawley"
                    /db_xref="taxon:10116"
                    /clone="UI-R-BU0-a02-h-08-0-UI"
                    /clone_1lb="UI-R-BU0"
                    /dev_stage="adult"
                    /lab_host="DH10B (Life Technologies)"
                    /note="Vector: pUT3D-Pac (Pharmacia) with a modified
                    polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BU0
                    library is a subtracted library derived from a mixture of
                    eye and ganglia tissues. For a detailed description of the
                    library from which this clone was derived, please visit
                    our web site at ratest.eng.uiowa.edu. The subtraction has
                    been previously described in (Bonaldo, Lennon and Soares,
                    Genome Research 6:791-806, 1996)
                    TAG_SEQ=None found"
BASE COUNT      67 a 142 c 140 g 109 t
ORIGIN
Query Match      64.8%; Score 386.6; DB 10; Length 458;
Best Local Similarity 90.4%; Pred. No. 5.3e-68;
Matches 413; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
OY 107 ACATCCCAACCATGAGGACCTTACCGGAGGTGATGAGCGGCGACAAGACGTGTGCA 166
      |||||||
Db 458 ACATCCCAACCATGAGGACCTTACCGGAGGTGATGAGCGGCGACAAGACGTGTGCA 399
OY 167 CGGTGACAGATACACAGACCAACCGGCGAGCCACCACTCCCGGCGATGAGCGCGCTGTCCA 226
      |||||||
Db 398 CACTGCAATCATCAGACGACCAACCGGCGAGCCACCACTCCCGGCGATGAGCGCGCTGTCCA 339

```

```

OY 227 TCTCCAAAGGGCAGCGCTTTCATCCGTCGTCGTCACCGACAAGAGTGTGGAG 286
      |||||||
Db 338 TCTCCAAAGGGCAGCGCTTTCATCCGTCGTCGTCACCGACAAGAGTGTGGAG 279
OY 287 AGCTGGGGCCCATCTTACCAAGCTCATCTGTCAGATCAAGGGCAGCGTGGAGGACATCCCG 346
      |||||||
Db 278 AGCTGAGCCCATCTTACCAAGCTCATCTGTCAGATCAAGGGCAGCGTGGAGGACATCCCG 219
OY 347 TGATCTCTGTGGGCAACAAGTGCATGAGAGCGACCGGAGGTGACACGCCGAGGGCG 406
      |||||||
Db 218 TCATGCTGTGTGGGCAACAAGTGTGATGAGACGCGACCGGAGGTGACACCGCGAGGGCG 159
OY 407 AGCGCGTGGCCCAAGAGGGAAGTGCCTTTTCATGAGACCGTCCGCGCAAGATGAATAC 466
      |||||||
Db 158 AGCGCGTGGCACAGAGAGTGCAGTGCCTTTCATGAGACCGTCCGCGCAAGATGAATAC 99
OY 467 ACGTCAGAGAGCTCTTCCAGAGAGTGTGACGCTGAGACGCGCGGGAACATGAGACTCA 526
      |||||||
Db 98 ACGTGAAGAGAGCTGTCCAGAGAGTGTGACGCTGAGACGCGCGGCGAGCTGACCTCA 39
OY 527 ACATGAGCGCAAGCGCTCCGCGGAGGACGAAGAGGAC 563
      |||||||
Db 38 GCGTGGACGGCAAGCGCTCCCAACAACATGAGCGGCGC 2

RESULT 7
BE095936/c 456 bp mRNA linear EST 12-JUN-2000
LOCUS      UI-R-BU0-apd-d-07-0-UI.s1 UI-R-BU0 Rattus norvegicus cDNA clone
DEFINITION UI-R-BU0-apd-d-07-0-UI 3', mRNA sequence.
ACCESSION  BE095936
VERSION     BE095936.1 GI:8486867
KEYWORDS    EST.
SOURCE      Norway rat.
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 456)
            Bonaldo,M.F., Lennon,G. and Soares,M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: mesoares@blue.weeg.uiowa.edu
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA-No.
FEATURES
    source          1..456
                    /organism="Rattus norvegicus"
                    /strain="Sprague-Dawley"
                    /db_xref="taxon:10116"
                    /clone="UI-R-BU0-apd-d-07-0-UI"
                    /clone_1lb="UI-R-BU0"
                    /dev_stage="adult"
                    /lab_host="DH10B (Life Technologies)"
                    /note="Vector: pUT3D-Pac (Pharmacia) with a modified
                    polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BU0
                    library is a subtracted library derived from a mixture of
                    eye and ganglia tissues. For a detailed description of the
                    library from which this clone was derived, please visit
                    our web site at ratest.eng.uiowa.edu. The subtraction has

```

been previously described in (Bonaldi, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_SEQ=None found"

Query Match 63.9%; Score 381.4; DB 10; Length 456;
Best Local Similarity 89.9%; Pred. No. 5.8e-67;
Matches 409; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

BASE COUNT 66 a 142 c 140 g 108 t
ORIGIN
QY 109 ATCCCAACCATTCAGGACCTTACCGGAGGATGATCAGTGGACGCAAGACCGTGTGACG 168
DB 456 ATCCCAACCATTCAGGACCTTACCGGAGGATGATCAGTGGACGCAAGACCGTGTGACG 397
QY 169 CTGACATCATCAGACACACCGGACGACGACGACGACGACGACGACGACGACGACGACG 228
DB 396 CTGACATCATCAGACACACCGGACGACGACGACGACGACGACGACGACGACGACGACG 337
QY 229 TCCAAAGGCGCAGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 288
DB 336 TCCAAAGGCGCAGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 277
QY 289 CTGAGGCGCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 348
DB 276 CTGAGGCGCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 217
QY 349 ATGCTGCTGGGCAACAGATGCGATGAGACGACGCGGAGGTGACACGCGGAGCGGACG 408
DB 216 ATGCTGCTGGGCAACAGATGCGATGAGACGACGCGGAGGTGACACGCGGAGCGGACG 157
QY 409 GCGGTGGCGCAAGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 468
DB 156 GCGGTGGCGCAAGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGG 97
QY 469 GTCAAGAGAGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 528
DB 96 GTCAAGAGAGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 37
QY 529 ATGACGCGCAAGCGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 563
DB 36 GTGAGCGGCAAGCGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2
RESULT 8
A1764459/c 422 bp mRNA linear EST 25-JUN-1999
LOCUS A1764459 422 bp mRNA linear EST 25-JUN-1999
DEFINITION UI-R-Y0-abj-f-09-0-01.s1 UI-R-Y0 Rattus norvegicus cDNA clone
ACCESSION A1764459
VERSION A1764459.1 GI:5210352
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 422)
AUTHORS Bonaldi,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-qt track not found, Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)

Seq primer: M13 Forward
FEATURES
source
POLYA-No.
Location/Qualifiers
1..422
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-abj-f-09-0-01"
/clone.lib="UI-R-Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
library is a subtracted library derived from an
individually-tagged normalized whole-eye (minus the lens)
library. The driver for the subtraction consisted of a
pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
UI-R-E1, UI-R-E0, and UI-R-C1). The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-qt track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-Y0) was constructed as follows: PCR
amplified cDNA inserts from previous library clones from
which 3' ESTs had been derived were used as a driver in a
hybridization with the normalized whole-eye library in
the form of single-stranded circles. The remaining
single-stranded circles (subtracted library) was purified
by hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-Y0
library. This procedure has been previously described
(Bonaldi, Lennon and Soares, Genome Research 6: 791-806,
1996)
TAG_LIB=UI-R-Y0
TAG_TISSUE=Eye
TAG_SEQ=CATTG"

BASE COUNT 63 a 133 c 128 g 98 t
ORIGIN
Query Match 59.3%; Score 353.8; DB 9; Length 422;
Best Local Similarity 90.0%; Pred. No. 2e-61;
Matches 379; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 143 TCAGCTGCGCAAGAGCGTGTGCAACGATCAGACACACCGGACGACGACGACG 202
DB 422 TCAGCTGCGCAAGAGCGTGTGCAACGATCAGACACACCGGACGACGACGACG 363
QY 203 TCCCGGCGATGCGCGCGTGTGCAACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 262
DB 362 TCCCGGCGATGCGCGCGTGTGCAACGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 303
QY 263 TCACAGCAAGAGTGTGCGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 322
DB 302 TCACAGCAAGAGTGTGCGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 243
QY 323 AGGGAGCGGTGAGGAGCATCTCCCGTATGCTCTGGGCAACAGTGCAGTAACCGCAGC 382
DB 242 AGGGAGCGGTGAGGAGCATCTCCCGTATGCTCTGGGCAACAGTGCAGTAACCGCAGC 183
QY 383 GGGAGGTGAGGAGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 442
DB 182 GGGAGGTGAGGAGCGGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 123
QY 443 AGACCTGCGCGCAAGATGAATACAGCTCAAGAGGCTTTCAGAGAGCTGTGACGCTGG 502
DB 122 AGACCTGCGCGCAAGATGAATACAGCTCAAGAGGCTTTCAGAGAGCTGTGACGCTGG 63
QY 503 AGACGCGCGCGCAAGATGAATACAGCTCAAGAGGCTTTCAGAGAGCTGTGACGCTGG 562
DB 62 AGACGCGCGCGCAAGATGAATACAGCTCAAGAGGCTTTCAGAGAGCTGTGACGCTGG 3
QY 563 C 563

Db 2 C 2

RESULT 9
AM533412/c 419 bp mRNA linear EST 06-MAR-2000
LOCUS
DEFINITION
UI-R-BU0-amb-f-04-0-UI.s1 UI-R-BU0 Rattus norvegicus CDNA clone
UI-R-BU0-amb-f-04-0-UI 3', mRNA sequence.
AM533412
ACCESSION
AM533412.1 GI:7175826
VERSION
KEYWORDS
EST.
SOURCE
ORGANISM
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 419)
REFERENCE
AUTHORS
TITLE
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
MEDLINE
COMMENT
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dt track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. CDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-No.

FEATURES
source
Location/Qualifiers
1..419
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BU0-amb-f-04-0-UI"
/clone_lib="UI-R-BU0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-BU0
library is a subcloned library derived from a mixture of
eye and ganglia tissues. For a detailed description of the
library from which this clone was derived, please visit
our web site at ratest.eng.uiowa.edu. The subtraction has
been previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG_LIB="UI-R-BU0"
TAG_TISSUE="ganglia"
TAG_SEQ="GCAGAC"

BASE COUNT 62 a 132 c 128 g 97 t

ORIGIN

Query Match 58.5%; Score 349.2; DB 10; Length 419;
Best Local Similarity 89.7%; Pred. No. 1,7e-60;
Matches 375; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Db 419 GCTGTGACAAAGCGGTGTCACGATCAGACACGCGGACGACCAAGTCC 205
|||||
GCTGTGACAAAGCGGTGTCACGATCAGACACGCGGACGACCAAGTCC 360

Qy 206 CGGCGATGACAGCGCTGTCATCTCCAGGCGACGCTTCATCTGCTTCCTCA 265
|||||
GCTGTGACAAAGCGGTGTCACGATCAGACACGCGGACGACCAAGTCC 360

Db 359 CAGCATGACAGCGGTGTCATCTCCAGGCGACGCTTCATCTGCTTCCTCA 300
|||||
GCTGTGACAAAGCGGTGTCACGATCAGACACGCGGACGACCAAGTCC 360

Qy 266 CCAGCAGCACTGCTGAGAGAGCTGGGCGCATCTACAGCTCATCTGAGATCAAG 325
|||||
GCTGTGACAAAGCGGTGTCACGATCAGACACGCGGACGACCAAGTCC 360

Db 299 CCAGCAGCACTGCTGAGAGAGCTGGGCGCATCTACAGCTCATCTGAGATCAAG 240
|||||
GCTGTGACAAAGCGGTGTCACGATCAGACACGCGGACGACCAAGTCC 360

Qy 326 GCAGCGTGAAGACATCCCGTGATGCTGTGGCAACAAGTGCATGAGACGACGGG 385
|||||
GCAGCGTGAAGACATCCCGTGATGCTGTGGCAACAAGTGCATGAGACGACGGG 180

Db 239 GCAGCGTGAAGACATCCCGTGATGCTGTGGCAACAAGTGCATGAGACGACGGG 180
|||||
GCAGCGTGAAGACATCCCGTGATGCTGTGGCAACAAGTGCATGAGACGACGGG 180

Qy 386 AGGTGACACGCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 445
|||||
AGGTGACACGCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120

Db 179 AGGTGACACGCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
|||||
AGGTGACACGCGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120

Qy 446 CCTGGCGCAAGATGAACATCAACGCTTCAGAGAGCTTCAGAGAGCTGAGAGAG 505
|||||
CCTGGCGCAAGATGAACATCAACGCTTCAGAGAGCTTCAGAGAGCTGAGAGAG 505

Db 119 CCTGGCGCAAGATGAACATCAACGCTTCAGAGAGCTTCAGAGAGCTGAGAGAG 60
|||||
CCTGGCGCAAGATGAACATCAACGCTTCAGAGAGCTTCAGAGAGCTGAGAGAG 60

Qy 506 CGCGCGGACATGAGAGCTTCACATGACGAGCGGCGGCGGCGGCGGCGGCGGCGG 563
|||||
CGCGCGGACATGAGAGCTTCACATGACGAGCGGCGGCGGCGGCGGCGGCGGCGG 563

Db 59 CGCGCGGACATGAGAGCTTCACATGACGAGCGGCGGCGGCGGCGGCGGCGGCGG 2
|||||
CGCGCGGACATGAGAGCTTCACATGACGAGCGGCGGCGGCGGCGGCGGCGGCGG 2

RESULT 10
BG702913 801 bp mRNA linear EST 07-MAY-2001
BG702913
LOCUS
DEFINITION
602684858F1 NIH_MGC_95 Homo sapiens CDNA clone IMAGE:4817247 5',
mRNA sequence.
ACCESSION
BG702913
VERSION
BG702913.1 GI:13974731
KEYWORDS
EST.
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 801)
REFERENCE
AUTHORS
TITLE
NIH-MGC <http://mhc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Miklos Palokvits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LRAM10717 row: d column: 16
High quality sequence stop: 760.

FEATURES
source
Location/Qualifiers
1..801
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4817247"
/clone_lib="NIH_MGC_95"
/lssue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: BluescriptR (modified
pBluescript KS+); Site 1: BamHI; Site 2: SalI-XhoI (gtcgag
); Oligo-dn primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size selected for average insert size 2.5 kb and
normalized to R07 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC library."

BASE COUNT 184 a 223 c 252 g 142 t

ORIGIN

Query Match 56.9%; Score 339.4; DB 12; Length 801;
Best Local Similarity 78.3%; Pred. No. 1,7e-58;
Matches 458; Conservative 0; Mismatches 121; Indels 6; Gaps 4;

Db 152 ATGCGTGAAGAGATTAACGCGGTGCTGTCGGGCGGCGGCGGCGGCGGCGGCAAG 60
|||||
ATGCGTGAAGAGATTAACGCGGTGCTGTCGGGCGGCGGCGGCGGCGGCGGCGGCAAG 211

OY	61	AGCTGCGTGTGTCCTGGCTTCGTGAAGGGACAGTTCCGGAGCACTTACCCTCCCAATC	120
Db	212	AGCTCCCTGGTGTTTAGGCTTTGTGAAGAAGCACATTCCGGAGAGACTACATCCCGACGGTG	271
OY	121	GAGGACACCTTACCGGCAGGTGATCAGCTGCGACAAGAGAGCTGTGCACGCTGAGATCACA	180
Db	272	GAAAGACACTTACCGGCCAAGTGTGATCAGCTGTGCAAGAGCATATGGACATTCGATCACC	331
OY	181	GACACACCGGAGGCACACAGTTCCCGGCATGACAGCGCTGTGC - ATCTCCAAGGGCCA	239
Db	332	GACACGAGGGGAGCCACACAGTTCCCGGCATGACAGCGCTGTGCATTCATCCAAAAGGGCA	391
OY	240	CGCCTTCATCCGTGTGTTCTCCGTCACAGCAGAAGCAAGTGTGAGAGAGACTGGGCCCCAT	299
Db	392	CGCCTTCATCCGTGTGTTCTCCGTCATTCACGACCGGACAGTCTTGGAGAGAGACTCAAGCCCAT	451
OY	300	CTACAGAGCTCATCGTGACAGATCAGGGCAGCGGTGAGAGCATCCCCTGATGCTGTGGG	359
Db	452	CTACCAACAATAATCTCGAGATCAAAGGGGAGCGCTGAGAGCATCCCATCATGCTGTGGG	511
OY	360	CACAAGTGGCGATGAGA --- CGCAGCGGAGGTGTGACACCGCGAGGCGCCAGGCGGTGGC	416
Db	512	GACACAGTGTGTGAGAGCCCGCAGCGGAGGTGTGAGAGCAGCAGAGGCGGAGGCGCTGTGGC	571
OY	417	CCAAAGAGTGGAGAGTCCGCTTTCATGAGACACTCGGCGCAAGATGAACTACAGTCAAGA	476
Db	572	CCGCACTATGGAAGTGTGCTTCATGTGAGAACCTCAGCCAAAGCT - CACCATTAAGCTGAAGA	630
OY	477	GCTCTTCCAGGAGCTGCT - GACGCTGTGAGAGCGCCCGCAATGAGACCTCAACATCGAGC	535
Db	631	GCTTTTTCAGGAGCTGCTCAAACTGTGGAAACCGCAGAGACCTGTCTCCAGATCGAGC	690
OY	536	GCAAGCGCTCCGGGAAGCAGAAGAGAGACAGACGCGGTCAAGGGCA	580
Db	691	GGAAAAAGACGACGACGCAAGAAAAGGACAGAGACTCAAGGGCA	735
RESULT 11			
BIS47693			
LOCUS	603191841P1 NIH_MGC_95 Homo sapiens CDNA clone IMAGE:5262995 5'	645 bp	mRNA linear EST 05-SEP-2001
DEFINITION	mRNA sequence.		
ACCESSION	BIS47693		
VERSION	BIS47693.1 GI:15435005		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 645)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cga@db-jemail.nih.gov Tissue procurement: Miklos Palkovits, M.D., Ph.D. CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki Toshuyuki and Piero Carninci (RIKEN) CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov plate: LHAM1663 row: a column: 12 High quality sequence stop: 634.		

FEATURES
Source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5262995"
/clone_lib="NH_MGC_95"
/tissue_type="hippocampus"

REFERENCE
AUTHORS
1 (pages 1 to 702)
Clark,M., Johnson,S.T., Lehnach,H., Lee,R., Li,F., Marra,M., Eddy
Eukaryota; Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Danio.

```

```

/lab host="DH10B"
/note="Organ: brain; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgagaa
); Oligo dt primed using primer 5'-TTTTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized for RoF 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHRRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT      161 a      166 c      205 g      113 t
ORIGIN

```

	Query Match	Similarity	59.4%;	Score 336.6;	DB 13;	Length 645;
	Best Local	Similarity	79.9%;	Pred. No. 6,1e-18;		
	Matches	434;	Conservative	0;	Mismatches 104;	Indels 5; Gaps 3
OY	1	ATGCCGAACAGATTAACGATTACCGCTGTTGTTTCGGGGGGGGCGGCTGGGCGAAG	60			
Db	75	ATGCGTGAACAGAGTAACGATTATCCGGGGGCGCGTTGTTGGGGCTGGCGATTGGCAGG	134			
OY	61	AGCTGCGTGTGCTCTGCGCTTCTGTAAGGGGACGTTTCCGGACACCTACATCCCAACG	120			
Db	135	AGCTCCCTGTGCTTTTGTAGGCTTTGTGTAAGGACATTTCCGGGAAGCTACATCCCGACGGTG	194			
OY	121	GAGGACACCTACCGGCGAGGTGATCAGCTTGCGACAGAGAGCGTGTGCACGCTGCAGATCACA	180			
Db	195	GAGAGACACCTACCGGCAAGTGTATCAGCTGTGACCAAGAGCATATGCACATTGCAGATCACCC	254			
OY	181	GACACCCACCGGACGACCCACGATGTTCCCGGCATATGCACAGCGCTGTCCATCTTCCAAGGGCCAC	240			
Db	255	GACACGAGCGGGGAGGACACCGATGTTCCCGGCATGACAGCGGCTGTCCATCTTCCAAGGGGCC	314			
OY	241	GCCCTCATCTGTTGGTTTTCGCTACACAGCAAGCAGTGCCTGGAGAGTGTGGGGCCCATC	300			
Db	315	GCCCTCATCTGTTGGTTTACTCCATTATCCAGCCGACAGTCTTGTGAGGAGCTCAAGCCCATC	374			
OY	301	TACAACTCATCTGTCAGATCAAGGGGACAGCGTTGAGAGACATCCCGGTGATGCTGTGGGC	360			
Db	375	TACGAACAAATCTGCGAGATCAAAAGGGGACGTGGAGAGCATCCCATCATGCTGTGGGGG	434			
OY	361	AACAGTCCGATGAGA---CGCAGGGGGAGGTGGAGACCCCGGAGGGCGGACGGGTGGGC	417			
Db	435	AACAAAGTGTGATGAGAGCCCCAGCCGCGAGGTGCAGAGCAGCGAGGGCGGAGGCTTTGGGC	494			
OY	418	CAAGAGTGAAGATGCGCTTTTCATGTAGAGACCTTGCGCAAGATGAACTTACAACTCAAGGAG	477			
Db	495	GCACA-TGGAAATGTGCCCTTCATGTGAGAACTGAGCCAAAGCTCAACCATATACCTGTAAGAG	553			
OY	478	CTCTTCCAGAGAGCTGCTGACGCTGAG-ACGCGCCGGAACATGAGCCTTACATCGACGG	536			
Db	554	CTTTTCCAGAGAGCTGCTCAACATGTGAGAAAGCGCAGAGACCGGTGAGTCTTCAAGATGACGG	613			
OY	537	CAA 539				
Db	614	GAA 616				

RESULT 12	AM305720	LOCUS	DEFINITION
	AM305720	702 bp	mRNA
	fj63a65.y1	zebrafish adult brain Danto	linear
	TR:095057	095057 BC41195_1..	EST 26-JUL-2002
			CDNA 5' similar to mRNA sequence.

ACCESSION	AW305720	GI:6718073
VERSION	AW305720.1	
KEYWORDS	EST	
SOURCE	zebrafish.	
ORGANISM	Danio rerio	

REFERENCE
AUTHORS

Euarxota, Metazoa; Chordata; Crinata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 702)
Clark, M., Johnson, S. L., Lehrbach, H., Lee, R., Li, F., Marra, M., Eddy

Db 254 CCACACGAGGAGGAGCCACCACTTCCCGCCATGCCGCGCTGTCCATAAGAGGC 313
 QY 239 AGCCCTTCATCTGTTGTTCTCCGTACACGACAGAGTGTGAGAGAGTGGGCCCA 298
 Db 314 AGCCCTTCATCTGTTGTTCTCCGTACACGAGAGTGTGAGAGAGTGGGCCCA 373
 QY 299 TCTACAGAGTGTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 357
 Db 374 TCTACAGAGTGTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 433
 QY 358 GGCACAGAGTGTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 413
 Db 434 GGCACAGAGTGTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 493
 QY 414 GGCACAGAGTGTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 473
 Db 494 GGCACAGAGTGTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 553
 QY 474 GGCACAGAGTGTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 533
 Db 554 GGCACAGAGTGTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 613
 QY 534 GGCACAGAGTGTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 590
 Db 614 GGCACAGAGTGTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTG 673
 QY 591 CATGTGA 597
 Db 674 CATGTGA 680

RESULT 14
 LOCUS BM172422 628 bp mRNA linear EST 04-DEC-2001
 DEFINITION Image4_2001/smu47bdf41.x1 NIH_MGC_95 Homo sapiens cDNA clone
 IMAGE:481673 5', mRNA sequence.
 ACCESSION BM172422
 VERSION BM172422.1 GI:17311985
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 628)
 AUTHORS Kale, P. I., Harsch, T. J., Folta, P. A., Nelson, D. O., Sanders, C. G. and
 Prange, C. K.
 TITLE The I.M.A.G.E. Consortium quality control effort: clone
 resequencing for verification
 JOURNAL Unpublished (2001)
 COMMENT The I.M.A.G.E. Consortium
 Lawrence Livermore National Laboratory
 Livermore, CA, USA
 Email: help@image.llnl.gov
 This read has been produced as part of the I.M.A.G.E. Consortium
 quality control effort. High quality sequence is defined as having
 100 or more base pairs with a phred quality value of 20 or greater,
 where a sliding window of 4 base pairs with a phred quality value
 of 15 or greater marks the beginning and end of the sequence. For
 information on obtaining this clone, please contact
 info@image.llnl.gov.
 Plate: LLM10715 row: 1 column: 18
 Seq primer: -21m13
 High quality sequence stop: 628.
 Location/Qualifiers
 1..628
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:481673"
 /clone_id="NIH_MGC_95"
 /tissue_type="hippocampus"
 /lab_host="DH10B"

FEATURES

source
 1..628
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:481673"
 /clone_id="NIH_MGC_95"
 /tissue_type="hippocampus"
 /lab_host="DH10B"

/note="Organ: brain; Vector: pBluescript (modified
 pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTNN-3',
 size-selected for average insert size 2.5 kb and
 normalized to R0F 5. This is a primary library enriched
 for full-length clones and constructed using the
 Cap-trapper method (Carlini, In preparation). Library
 constructed by M. Brownstein (NIH/NHGRI, National
 Institutes of Health). Note: this is a NIH_MGC Library."
 BASE COUNT 139 a 175 c 197 g 116 t 1 others
 ORIGIN

Query Match 52.3%; Score 312.2; DB 13; Length 628;
 Best Local Similarity 79.3%; Pred. No. 4.9e-53;
 Matches 395; Conservative 0; Mismatches 99; Indels 4; Gaps 2;

QY 1 ATCCGGAACAGATGATACGATGATGATGATGATGATGATGATGATGATGATG 60
 Db 130 ATCCGGAACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 189
 QY 61 AGCTCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
 Db 190 AGCTCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 249
 QY 121 GAGACACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTAC 180
 Db 250 GAGACACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTACCTAC 309
 QY 181 GACACACCGGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 240
 Db 310 GACACACCGGAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAGCCAG 369
 QY 241 GCCTTCATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 300
 Db 370 GCCTTCATCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 429
 QY 301 TACAACTATCTGTAAGTACAACTGTAAGTACAACTGTAAGTACAACTGTAAG 359
 Db 430 TACAACTATCTGTAAGTACAACTGTAAGTACAACTGTAAGTACAACTGTAAG 489
 QY 360 CACAACTGTAAGTACAACTGTAAGTACAACTGTAAGTACAACTGTAAGTACAA 416
 Db 490 GAACTGTAAGTACAACTGTAAGTACAACTGTAAGTACAACTGTAAGTACAACTG 549
 QY 417 CCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 476
 Db 550 CCAAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 609
 QY 477 GCCTTCAGAGAGCTGCT 494
 Db 610 AGCTTCAGAGAGCTGCT 627

RESULT 15
 LOCUS BG699614 709 bp mRNA linear EST 07-MAY-2001
 DEFINITION 602679344F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:4811962 5',
 mRNA sequence.
 ACCESSION BG699614
 VERSION BG699614.1 GI:13968106
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 709)
 NIH-MGC http://mgc.ncl.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs@mail.nih.gov
 Tissue Procurement: Miklos Palokvits, M.D., Ph.D.
 cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 15:14:42 ; Search time 157.566 Seconds
(Without alignments)
2938.192 Million cell updates/sec

Title: US-09-873-546-4

Perfect score: 597
Sequence: 1 atgcgcgaacagagtaacga.....gcaatgcacccatcgtga 597

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 53826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCR_NEM_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCR_NEM_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	234.4	39.3	452	10	US-09-960-352-4253
2	204.6	34.3	454	10	US-09-960-352-5830
3	151	25.3	3082	10	US-09-778-963A-1
4	137.8	23.1	551	10	US-09-765-298A-25
5	137.8	23.1	570	9	US-09-359-595-7
6	137.8	23.1	570	10	US-09-765-298A-27
7	136.2	22.8	570	12	US-10-104-484-1
8	136.2	22.8	570	12	US-10-104-484-3
9	118.2	19.8	3346	9	US-09-764-868-67
10	118	19.8	688	9	US-09-764-868-490
11	107.8	18.1	1119	9	US-10-067-813-1
12	107.8	17.2	1105	9	US-10-067-813-3
13	100.4	16.8	11221	10	US-09-778-963A-3
14	99	16.6	405	10	US-09-960-352-10273
15	97	16.2	368	10	US-09-864-761-21643
16	84.2	14.1	612	10	US-09-972-529-3
17	84.2	14.1	1224	10	US-09-972-529-1
18	84.2	14.1	3192	10	US-09-788-654A-1
19	81.2	13.6	3936	10	US-09-919-172-49

20	80.8	13.5	752	9	US-09-764-868-501	Sequence 501, App
21	80.4	13.5	1271	12	US-10-044-090-110	Sequence 110, App
22	72.4	12.1	925	10	US-09-967-736-4	Sequence 4, App1
23	72	12.1	969	10	US-09-801-368-287	Sequence 287, App
24	67	11.2	407	10	US-09-960-352-11732	Sequence 11732, A
25	67	11.2	267	10	US-09-925-300-678	Sequence 678, App
26	66.4	11.1	942	10	US-09-962-832-256	Sequence 256, App
27	66.2	11.1	2021	9	US-09-764-868-88	Sequence 88, App1
28	66.2	11.1	3257	10	US-09-817-198A-1	Sequence 1, App1
29	65.6	11.0	4083	10	US-09-817-182-1	Sequence 1, App1
30	65	10.9	952	10	US-09-917-800A-1389	Sequence 1389, App
31	64	10.7	930	10	US-09-801-368-285	Sequence 285, App
32	63.8	10.7	2183	10	US-09-917-800A-1104	Sequence 1704, App
33	63.6	10.7	1191	12	US-10-044-090-111	Sequence 111, App
34	61.2	10.3	2291	9	US-10-098-841-237	Sequence 84, App1
35	61.2	10.3	2418	9	US-09-764-868-84	Sequence 84, App1
36	61.2	10.3	2479	9	US-09-764-868-69	Sequence 69, App1
37	60.8	10.2	1157	9	US-10-108-605-44	Sequence 44, App1
38	60.6	10.2	1443	10	US-09-962-832-224	Sequence 224, App
39	60.2	10.1	566	9	US-09-764-868-507	Sequence 507, App
40	59.6	10.0	4017	10	US-09-880-107-3029	Sequence 3029, App
41	57.2	9.6	1007	10	US-09-822-849A-563	Sequence 563, App
42	57	9.5	1274	10	US-09-925-302-91	Sequence 91, App1
43	56.4	9.4	385	10	US-09-960-352-13143	Sequence 13143, A
44	56.4	9.4	1077	9	US-09-764-868-497	Sequence 497, App
45	56.4	9.4	1364	9	US-09-764-868-77	Sequence 77, App1

ALIGNMENTS

RESULT 1
US-09-960-352-4253
Sequence 4253, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengping
APPLICANT: Bvatt, John C.
APPLICANT: Mathiaslagn, Nagappan
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960.352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 4253
LENGTH: 452
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 18-LIB34-086-Q1-E1-E5
US-09-960-352-4253

Query Match 39.3%; Score 234.4; DB 10; Length 452;
Best Local Similarity 71.5%; Pred.No. 2, le-43;
Matches 323; Conservative 0; Mismatches 126; Indels 3; Gaps 1;
Db 1 GTGATCGGCTCGGCGGCGGTGGCAAGCGCTGTGTCGAAGGTGGCGGCAC 60
OY 34 GTGTTGGGGGGGGGGGGGGGCAAGAGCTCGTGTGTCGCTTGTGTAAGGCACG 93
OY 94 TTCCCGACCTTACATATCCCAAGAGAGCTTACCGGAGGTATGATCGGCAC 153
Db 61 TTCCGTGAGGCTTACCTCCAGCAGATACCTACCGGAGGCGCTAGCTGACAG 120
OY 154 AAGAGCGTGTGACAGCTGCAGATACAGACACCGGAGCGACCACTTCCGGCCATG 213
Db 121 CACAGGCGGGGTCCTACATACAGACACCGGAGGCGGCGCTACCGGCGCTG 180
OY 214 CAGCGCTTGCATTCATCCAGAGGCGCAGGCTTATCTGTTTCTCCGTACCGCAAG 273
Db 181 CAGCGCTTGCATTCATCCAGAGGCGGCTTATCTGTTTATTCATACCGAGAG 240

QY 379 CAGCGGAGTGGACACGCGCGAGCGGTCGCCCAAGTGAAGTGCCTTC 438
11 111111 11 111111 111111 111111 111111 111111
DB 364 GCACGACGCTGTGGAAATCTGGCAGGCTCAGACCTGCCGGAAGCTACGCCCTTAC 423
QY 439 ATGAGACCTCGCGCAAGATGAATCAACGTCACAGGACCTTTCAGAGACTGCTG 495
11 1111111111111111 11 1111 11111 11111 11111
DB 424 ATCGAGACCTCGCGCAAGACCGGAGGAGTGGAGATGCTTTCACAGCTTGGTG 480

RESULT 9

US-09-764-868-67
; Sequence 67, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 67
; LENGTH: 3346
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (2787)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (2795)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-67

Query Match

Best Local Similarity 19.8%; Score 118.2; DB 9; Length 3346;
Matches 258; Conservative 0; Mismatches 233; Indels 0; Gaps 0;

QY 9 ACAGAGTAACGATTAACCGGCTGTGTTCCGGGCGGCGGCGTGGCAAGCTCGCT 68
11 11 1111 11 11111111 11 11 111111 11 111111 11
DB 188 ACTCATGAGGGAATACAGATGAGTGGTGTAGGAGTGGAGGCTTGCACCAATCTGCCCT 247
QY 69 GGTGTCGCTTCGTGAAGGACGTTCCGCGACCTACATCCACCATGAGGACAC 128
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 248 TACTGTGACGTTTGTACTGAGGCTTTCATTTGAAATATGACCCCATTTGAAAGTTT 307
QY 129 CTACCGGAGGTGATCAGCTGCGACAAGAGCTGTGACGCTGCAGATCAGACACAC 188
111111 111111 1111 111 111111 11 111111 11 111111 11
DB 308 CTACCGCAAGAGATGAGTGAAGTGAATCTTCCCTCCGCTGCGAATTTCTGAGACCGC 367
QY 189 CGGACGACCAAGTTCGCGGCATGCAAGCGCTGTCCATCTCCAGAGCCACGCTTCAT 248
11 11 111111 11111 11111 11 1111 11 11111 11111 11111
DB 368 AGAAGCTGAGCAAGTTCCTCCATGAGAGATCTCTACATCAAAAAGCGCAAGTTTCAT 427
QY 249 CCGTGTCTTCGTCACGACGAGGAGCTGCGAGGAGCTGGGGCCATCTACAGCT 308
111111 111 11111 11111 11111 11111 11111 11111 11111 11111
DB 428 CCGTGTCTTCGTCACGACGAGGAGCTGCGAGGAGCTGGGGCCATCTACAGCT 308
111111 111 11111 11111 11111 11111 11111 11111 11111 11111
QY 309 CATCTGTCAGATCAAGGAGCGTGGAGACATCCCGTATGCTCTGTGGCAACAAGT 368
111 111 1111 1111 1111 111 111 111 111 111 111 111 111 111
DB 488 AATTTGTCAGATGAAGATGATGAAGAGTCCCATTAATCTGTAGAAATTAAGTGA 547
QY 369 CGATGAGACGACGCGGAGTGGACACGCGGAGGCGTGGTGGCCCAAGATGGA 428
11 11 111111111111 11 11 1111 11111 11111 11111 11111
DB 548 TCTGGAACCAAGAGAGTATGCTTTCAGAGGACAGAGCTCTGCTCAAAATGCGG 607
QY 429 GTGCGCTTTCATGAGACCTCGGCAAGATGAATCAACGTCACAGAGCTTTCAGGA 488
11 11111111111111 11111 111 111 111 111 111 111 111 111
DB 608 CTGTCTTTCATGAGACATCGCAAAAGTAAATCAATGTGATGAATTTTCTGA 667
QY 489 GCTGTGACGC 499

DB 668 GATCTGACGC 678

RESULT 10

US-09-764-868-490
; Sequence 490, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 490
; LENGTH: 688
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (579)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (610)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (669)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-490

Query Match

Best Local Similarity 19.8%; Score 118; DB 9; Length 688;
Matches 256; Conservative 1; Mismatches 234; Indels 0; Gaps 0;

QY 9 ACAGAGTAACGATTAACCGGCTGTGTTCCGGGCGGCGGCGTGGCAAGAGCTCGCT 68
11 11 1111 11 11111111 11 11 111111 11 11 111111 11 -111
DB 181 ACTCATGAGGGAATACAGATGAGTGGTGTAGGAGTGGAGGCTTGCACCAATCTGCCCT 240
QY 69 GGTGTCGCTTCGTGAAGGACGTTCCGCGACCTACATCCGCAAGTGAAGACAC 128
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 241 TACTGTGACGTTTGTACTGAGGCTTTCATTTGAAATATGACCCCATTTGAAGATT 300
QY 129 CTACCGGAGGTATCAGCTGCGACAAGAGCTGTGACGCTGCAGATCAGACACAC 188
111111 111111 1111 111 111111 11 111111 11 111111 11
DB 301 CTACCGCAAGAGATGAGTGAAGTGAATCTTCCCTCCGCTGCGAATTTCTGAGACACCGC 360
QY 189 CGGACGACCAAGTTCGCGGCATGCAAGCGCTGTCCATCTCCAGAGCCACGCTTCAT 248
111111 111111 11111 11111 11 1111 11 11111 11 11111 11
DB 361 AGAAGCTGAGCAAGTTCCTCCATGAGAGATCTCTACATCAAAAAGCGCAAGTTTCAT 420
QY 249 CCGTGTCTTCGTCACGACGAGGAGCTGCGAGGAGCTGGGGCCATCTACAGCT 308
111111 11 11111 11111 11111 1111 1111 11 1111 11 1111 11
DB 421 CCGTGTCTTCGTCACGAGCTGTTAATCAACAGCTTTCATGAGATATCAAGCAATGAGATCA 480
QY 309 CATCTGTCAGATCAAGGAGCGTGGAGGATCCCGTATGCTCTGTGGCAACAAGT 368
11 11 1111 1111 1111 11 11 111 111 111 111 111 111 111
DB 481 AATTTGTCAGATGAAGATGATGAAGAGTCCCATTAATCTCKAGTGAAGAAATTAAGTGA 540
QY 369 CGATGAGACGACGCGGAGTGGACACGCGGAGGCGGAGGCGTGGCCCAAGATGGA 428
11 11 111111 11111 111 111 111 111 111 111 111 111 111
DB 541 TCTGGAACCAAGAGAGTATGCTTTCAGAGGACAGAGCTCTGCTCAAAATGCGG 600
QY 429 GTGCGCTTTCATGAGACCTCGGCAAGATGAATCAACGTCACAGAGCTTTCAGGA 488
11 11111111111111 1111 11 11 11 11 11 11 11 11 11 11
DB 601 CTGTCTTTCATGAGACATCGCAAAAGTAAATCAATGTGATGAATTTTCTGA 660
QY 489 GCTGTGACGC 499
DB 661 GATCTGACGC 671

```
? ? ID NO 1  
? LENGTH: 1119  
? TYPE: DNA  
? ORGANISM: Murinae gen. sp.  
? FEATURE:  
? NAME/KEY: CDS  
? LOCATION: (19)..(642)  
?-S-10-067-813.-1
```

OY	29	TGTTGGCTTTCGGGGCGGGCCGTGGGGCAAGACCTGCCTGTGCTGCCTTCGTGAAG	88
Db	65	TACTATGTGGTGGAATGTGTGTGGCAGAAGTGCCTCATATTCAGTTTTTCCAGA	122
OY	89	GCACTTCCGGACACCACATATCCCACATCAGAGACACCTACCGAGAGTGA	144
Db	125	AGATCTTTTGCTGCCTACTACGACCCCCACATTTGAAGATCTTACTCTGAAGATCACAGA	184
OY	149	GCACAAGAGCGTTGTGACGCTGCAGATCAACAGACACACGGCAGCACCAGTATCCCG	208
Db	185	TTTGCAATCACTAGTGGGCCATCTTGGATGTTCAGACACAGCGGGCAGSAGSAGTTCA	244
OY	209	CCATCGAGCGGCTGTCCATCTCCAAGGGCCACGCCCTTATCTGTGTTTCCGTACCA	268
Db	245	CCATCGCGGAACANTATACATGCGCCAAAGGGATGGCTTCTCATTTGTACTCCCTCACG	304
OY	269	GCAAGCAGTGCCTGGAGAGACTGGGGCCCATCTACAAGCTCATGCTCAGATCAAAGGCA	328
Db	305	ACAAGGCCACTTCGAGCAGCTGGATCCGCTTCACACAGCTCATTTCTCGTTCAGAGCA	364
OY	329	GCGTAGAGAACATCCCGGTATGCTGTGGGCAACAAGTGCATGAGACCGAAGC--GGG	388
Db	365	G---GGAGTCATTCCCATATGATCTCTGTGGCCAAAGGGTATCTGATCACCTTAAGGA	424
OY	386	AGGTGACACGCGCGAGGCGCAGGCGGTGGCCCAAGATGTGAAGTGGCTTCATGTAGGA	448
Db	422	AAGTCACCAAGGAGCAAAAGAAAATGTGCACACCAATATCATATTCCTATATTAAGGA	484
OY	446	CCCTCGGCGAAG 456	
Db	482	CCAGTSCCAAG 492	

RESULT 12
US-10-067-813-3
Sequence 3, Application US/10067813
Patent No. US20020156013A1
GENERAL INFORMATION:
APPLICANT: Renault, Jean-Christophe
APPLICANT: Louahed, Jamila

Query Match	17.2%;	Score 102.6;	DB 9;	Length 1085;
Best Local Similarity	56.2%;	Pred. No. 3.9e-14;		
Matches 214; Conservative	0;	Mismatches 164;	Indels 3;	Gaps 1

QY	22	TACCGCTGGTGGTGTGTTGGGGGCGGGCGGTGGGGCAAGAGCTGGTGGTCCGCTTC	81
Db	141	TACAAAGCTGGTGGTGGTGGGGATGGGGGTGTGGGCAAAATGGCTTCACCATCCAGTTT	200
QY	82	GTGAAGGGCAGCTTCCCGGACACCTTACATCCCAACCATCGAGAGACCTTACCGGACAGTG	141
Db	201	TTCCAGAAAGATCTTTGGTCCGTGACTATGAGCCCAACATTTGAAGACTCTTACCTTAAACAT	266
QY	142	ATCAGCTGGCAGCAAGAGCGGTGTGCACGCTGCAGATCAACAGACACCAACCGGACGACACAG	201
Db	261	ACGGAGATTGGACATCAATGAGGCGCATCTTGGACGTTCTGACACACGCTGGGACAGAGAA	320
QY	202	TTCCCGGCGATGAGCGCCGCTGTCCATCTCCAAAGGGCCACGCTTATCTGGTGTTC	261
Db	321	TTTACGCCCATGGCGGGAGCAATACATCGCACGGGGATGGCTTCCATCTGTCTTACTTC	380
QY	262	GTCCACCAAGCAGTAGTGGTGGAGAGACTGGGGGCCATATCAAGAGCTCATGTGGACATC	321
Db	381	GTCACTGACAAAGCCACACTTTTGACACAGTGGACCGCTTCCACCAACGCTTATCTTGGCGTTC	440
QY	322	AAGGGCAGCGTGAAGGACATCCCGTGTACTGTGGGCAACAGTGGATGAGACAG	381
Db	441	AAAGACAG--GGAGTCAATTCGCCGATATCTCTGTGGCAACAAAGTGCAATTTGATGACAC	497
QY	382	CGGAGGTTGGACACGCGCCGAG	402
Db	498	TTTGAAGGAATCAACGAGGAG	518

```

RESULT 13
US-09-778-963A-3
; Sequence 3, Application US/09778963A
; Patent No. US200201172A1
GENERAL INFORMATION:
; APPLICANT: NEELEAM, Beena et al
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: C1001112
; CURRENT APPLICATION NUMBER: US/09/778,963A
; CURRENT FILING DATE: 2001-02-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 11221
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-778-963A-3

```


OTHER INFORMATION: SWISSPROT HIT: P13856, EVALU 8.00e-12
OTHER INFORMATION: EST_HUMAN HIT: BE38944.1, EVALU 0.00e+00
US-09-864-761-21643

Query Match 16.2%; Score 97; DB 10; Length 368;
Best Local Similarity 65.4%; Pred. No. 5.7e-13;
Matches 142; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 16 AACGATTACCGCGTGTGTGTCGGGGCGGCGCGTGGGCAAGAGCTGCTGTGCTG 75
DB 149 AACTCATACCGCATGTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 208
QY 76 CGCTTCGTGAAGGGCAAGTTCGCGACACCTACATCCCAACATCGAGGACACTACCG 135
DB 209 CGCTTCCTCATATGGCCGCTTTGAGACCAATACACCCACATCGAGGACTTCACCGT 268
QY 136 CAGGTGATCAGTGCAGACAAGAGCTGTGCACGCTGCAGATCACAGACACACCGGACG 195
DB 269 AAGGTATACACATCCGCGCGGCGGACATGTACACAGCTGCACATCTCTGATACCTTGCGAAC 328
QY 196 CACCAAGTCCCGGCGCATGCGAGCGCGCTGTCCATCTCA 232
DB 329 CACCCCTTCGCGCATGCGAGGCTGTCCATCTCA 365

Search completed: March 23, 2003, 17:17:10
Job time : 166.566 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:56 ; Search time 106.909 Seconds

(without alignments)
8438.824 Million cell updates/sec

File: US-09-873-546-3

Perfect score: 31

Sequence: 1 gcggaattccatgaggggtcattgcc 31

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Genbmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_ov:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rtd:*
- 36: em_htg_man:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	23	74.2	774	9	AB076888 Homo sapi
C 2	23	74.2	1249	6	AX430295 Sequence
C 3	23	74.2	2827	9	AK096600 Homo sapi
C 4	23	74.2	3061	6	AX430418 Sequence
C 5	23	74.2	3391	9	BC030660 Homo sapi
C 6	23	74.2	177540	9	AC006538 Homo sapi
C 7	23	74.2	194456	2	AC113495 Mus muscu
C 8	22.8	73.5	153873	9	AC026458 Homo sapi
C 9	22	71.0	597	9	AY056037 Homo sapi
C 10	22	71.0	597	9	AY056041 Homo sapi
C 11	21.2	68.4	67767	2	AC068480 Homo sapi
C 12	21.2	68.4	185829	2	AC127182 Rattus no
C 13	21	67.7	3255	8	AF348621 Emericell
C 14	21	67.7	137615	2	AP005284 Oryza sat
C 15	20.6	66.5	15083	2	AC112349 Rattus no
C 16	20.6	66.5	169264	2	AC112889 Rattus no
C 17	20.6	66.5	191557	9	AC008506 Homo sapi
C 18	20.4	65.8	151283	2	AC102763 Mus muscu
C 19	20.4	65.8	152470	9	AL135923 Human DNA
C 20	20.4	65.8	157692	10	AL672177 Mouse DNA
C 21	20.4	65.8	174704	2	AC122577 Rattus no
C 22	20.4	65.8	186989	2	AC101913 Mus muscu
C 23	20.4	65.8	219730	2	AC116661 Mus muscu
C 24	20.2	65.2	158450	2	AC068082 Homo sapi
C 25	20.2	65.2	162419	2	AC093737 Homo sapi
C 26	20.2	65.2	192396	9	AC012668 Homo sapi
C 27	20.2	65.2	347162	10	AC084070 Mus muscu
C 28	20	64.5	25100	3	U49946 Caenorhabdi
C 29	20	64.5	68206	2	AC124323 Mus muscu
C 30	20	64.5	70527	2	AC121560 Mus muscu
C 31	20	64.5	110000	2	AC013622_3 Continuation (4 of
C 32	20	64.5	138783	9	HS564M11 Human DNA
C 33	20	64.5	142422	9	AC113935 Homo sapi
C 34	20	64.5	146072	2	AC091233 Oryza sat
C 35	20	64.5	158145	2	AC128627 Rattus no
C 36	20	64.5	161179	2	AC066601 Homo sapi
C 37	20	64.5	171296	9	AL391262 Human chr
C 38	20	64.5	173663	2	AC096960 Rattus no
C 39	20	64.5	180366	9	AC036196 Homo sapi
C 40	20	64.5	189540	10	AL772264 Mouse DNA
C 41	20	64.5	193121	2	AC119482 Rattus no
C 42	20	64.5	193865	2	AL845164 Mus muscu
C 43	20	64.5	194237	9	AC068870 Homo sapi
C 44	20	64.5	194487	2	AC127043 Rattus no
C 45	20	64.5	194492	9	AC097625 Homo sapi

ALIGNMENTS

RESULT 1	AB076888	774 bp	mrna	linear	PRI 25-JUN-2002
LOCUS	AB076888				
DEFINITION	Homo sapiens mRNA for Di-Ras1, complete cds.				
ACCESSION	AB076888				
VERSION	AB076888.1	GI:21624247			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens adult brain cDNA to mRNA, clone_11b:lambda ZAPII human brain cDNA.				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	Kontani,K., Tada,M., Ogawa,T., Okai,T. and Katada,T.				
TITLE	Di-Ras: A Distinct Subgroup of Ras-Family GTPases with Unique				

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Biochemical Properties
Unpublished
2 (bases 1 to 774)
Kontani,K., Ogawa,T., Okai,T., Tada,M. and Katada,T.
Direct Submission
Submitted (24-DEC-2001) Toshiaki Katada, University of Tokyo,
Department of Physiological Chemistry, Graduate School of
Pharmaceutical Sciences; 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033,
Japan (E-mail:katada@mol.f.u-tokyo.ac.jp, Tel:81-3-5841-4750,
Fax:81-3-5841-4751)
Location/Qualifiers
1..774
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13.3"
/tissue-type="brain"
/clone_lib="lambda ZAPII human brain cDNA"
/dev_stage="adult"
1..774
/gene="Di-ras1"
100..696
/gene="Di-ras1"
/note="member of the Ras family
small GTP-binding protein"
/codon_start=1
/product="Di-Ras1"
/protein_id="BAC01115.1"
/db_xref="GI:21624248"
/translation="MPESNDYRVAVFGAGVGKSSLYLRFVKGFRPTVPTIEDTY
RQVISCBSVCTLTDTGDSHOPAPMORLSISGHAFILVFSYTSKSLDELPIYK
LIVQKSGVEDIPMYLGNKCDEREDTREAQVADQEMKCAPMETSAKMNVKEL
FQELTLERFNMSLINDGKRSQKQRTDRVKGCTLM"

BASE COUNT
154 a 268 c 239 g 113 t

ORIGIN
Query Match
Best Local Similarity 100.0%; Score 23; DB 9; Length 774;
Pred. No. 2.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 CTCACATGAGGGTGCAATTGGCC 31
|||||
Db 697 CTCACATGAGGGTGCAATTGGCC 675

RESULT 2
AX430295/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
ORGANISM
SOURCE
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Chalup,M.S., Altus,C.M., Lincoln,S.E., Dufour,G.E. and Jackson,S.
Molecules for disease detection and treatment
Patent: WO 0240715-A 77 23-MAY-2002;
INCYTE GENOMICS INC (US)
Location/Qualifiers
1..1249
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: LG:389870.1:2000SEP08"

BASE COUNT
230 a 436 c 355 g 228 t

ORIGIN
Query Match
Best Local Similarity 74.2%; Score 23; DB 6; Length 1249;
Pred. No. 2.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 CTCACATGAGGGTGCAATTGGCC 31
|||||
Db 647 CTCACATGAGGGTGCAATTGGCC 625

RESULT 3
AK096600/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan: cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers
1..2827
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="OCBBF2011067"
/tissue-type="brain"
/clone_lib="OCBBF2"
/dev_stage="fetal"
/note="cloning vector: PME18SFL3"

BASE COUNT
688 a 750 c 983 g 406 t

ORIGIN
Query Match
Best Local Similarity 74.2%; Score 23; DB 9; Length 2827;
Pred. No. 2.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 CTCACATGAGGGTGCAATTGGCC 31
|||||
Db 727 CTCACATGAGGGTGCAATTGGCC 705

RESULT 4
AX430418/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Makamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Nagatsuna,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2827)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan: cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
Location/Qualifiers
1..3061
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="OCBBF2011067"
/tissue-type="brain"
/clone_lib="OCBBF2"
/dev_stage="fetal"
/note="cloning vector: PME18SFL3"

BASE COUNT
3061 a 3061 c 3061 g 3061 t

ORIGIN
Query Match
Best Local Similarity 74.2%; Score 23; DB 9; Length 3061;
Pred. No. 2.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 CTCACATGAGGGTGCAATTGGCC 31
|||||
Db 727 CTCACATGAGGGTGCAATTGGCC 705

BASE COUNT	650 a	996 c	1016 g	729 t
ORIGIN				
Query Match	74.2%	Score 23;	DB 9;	Length 3391;
Best Local Similarity	100.0%;	Pred. No. 2, 3;		
Matches 23;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
OY	9	CTCACATGAGGGTGCGATTGGCC 31		
Db	735	CTCACATGAGGGTGCGATTGGCC 713		
RESULT 6				
LOCUS	AC006538	177540 bp	DNA	linear
DEFINITION	Homo sapiens chromosome 19, BAC 41195 (CIT-B-31c16), complete			
ACCESSION	AC006538			
VERSION	AC006538.1	GI:4235145		
KEYWORDS	HTG.			
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Viswanathan,V., Burkhart-Schultz,K., Gordon,L., Dias,J., Ramirez,M., Stillwegen,S., Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Garnes,J., Dangnan,L., Erlar,A., Christensen,M., Georgescu,A., Avila,J., Liu,S., Atlix,C., Andreise,T., Traubheim,M., Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Krommiller,B., Arellano,A., Sanders,C., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.			
TITLE	Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 177540)			
AUTHORS	Lamerdin,J.E.			
TITLE	Direct Submission			
JOURNAL	Submitted (07-FEB-1999) Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA			
COMMENT	Map and sequence oriented from p telomere to centromere. BC41195 (CIT-B-31c16) is currently separated from cosmid R32203 (AC006275) to the left by a sequence gap of approximately 6 kb, and overlaps BAC 102889 (CIT-B-191n6; AC006130) to the right from bases 175,545 to 177,540. Additional map and sequence information are available at: http://www.jgi.llnl.gov/bdip/genome/genome.html .			
FEATURES	Location/Qualifiers			
SOURCE	1. 177540			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/chromosome="19"			
	/map="19p13.3 between CDC34 and D19S342"			
	/clone="CIT-B-31c16"			
	/clone_11b="Cal tech CIT-B BAC library"			
	/note="11b, clone name: BC41195"			
	complement(1..144)			
	/rpt_family="LIMC/D"			
	283..381			
	/rpt_family="AluJo"			
	complement(399..437)			
	/rpt_family="(GAA)n"			
	complement(438..738)			
	/rpt_family="AluXx"			
	complement(866..931)			
	/rpt_family="LIMB7"			

```
misc_feature      1175..1365  
/note="predicted exon, program: grail2exons_human_1.3,  
frame: 2, quality: good, score: 57.000"  
repeat_region    /rpt_family="AluSg"  
complement(2133..2231)  
repeat_region    /rpt_family="MER81"  
2739..2935  
repeat_region    /rpt_family="AluJo"  
2946..3247  
repeat_region    /rpt_family="AluSx"  
3256..3384  
repeat_region    /rpt_family="L1MC3"  
3389..3604  
repeat_region    /rpt_family="AluJo"  
3631..3716  
repeat_region    /rpt_family="L1MD3"  
complement(4087..4242)  
repeat_region    /rpt_family="AluJo"  
complement(4294..4436)  
repeat_region    /rpt_family="AluJo"  
complement(4598..4887)  
repeat_region    /rpt_family="AluSg"  
complement(4906..5373)  
repeat_region    /rpt_family="L1MB7"  
complement(5382..5669)  
repeat_region    /rpt_family="AluSx"  
complement(6425..6609)  
repeat_region    /rpt_family="(GGA)n"  
complement(6610..6765)  
repeat_region    /rpt_family="(GAA)n"  
complement(6766..7065)  
repeat_region    /rpt_family="AluSx"  
7093..7392  
repeat_region    /rpt_family="AluSx"  
7423..7511  
repeat_region    /rpt_family="L1MA7"  
7539..7839  
repeat_region    /rpt_family="AluSx"  
7852..7938  
repeat_region    /rpt_family="L1MA7"  
8001..8312  
repeat_region    /rpt_family="AluJo"  
complement(8771..8897)  
repeat_region    /rpt_family="MER20"  
complement(9532..9722)  
repeat_region    /rpt_family="MER20"  
complement(9970..10143)  
repeat_region    /rpt_family="MER20"  
10144..10405  
repeat_region    /rpt_family="AluJo"  
10421..10630  
repeat_region    /rpt_family="L1M3"  
10631..10756  
repeat_region    /rpt_family="AluSg"  
10757..11058  
repeat_region    /rpt_family="AluY"  
11068..11243  
repeat_region    /rpt_family="AluSg"  
11247..11520  
repeat_region    /rpt_family="L1M3"  
11522..11816  
repeat_region    /rpt_family="AluSg"  
complement(11827..12033)  
repeat_region    /rpt_family="MER58A"  
12255..12555  
repeat_region    /rpt_family="AluJo"  
complement(12562..12617)  
repeat_region    /rpt_family="(TA)n"  
12769..12894  
repeat_region    /rpt_family="FLAM_A"  
complement(12901..12951)  
repeat_region    /rpt_family="L1MC3"  
  
repeat_region    12960..13013  
repeat_region    /rpt_family="LFR13"  
13013..13408  
repeat_region    /rpt_family="LFR13"  
complement(13409..13457)  
repeat_region    /rpt_family="L1MD3"  
13651..13908  
repeat_region    /rpt_family="L1MB8"  
13906..14116  
repeat_region    /rpt_family="L1MC2"  
complement(14173..14468)  
repeat_region    /rpt_family="AluY"  
14497..14811  
repeat_region    /rpt_family="AluSx"  
14812..15011  
repeat_region    /rpt_family="(GGA)n"  
15689..16007  
repeat_region    /rpt_family="AluJo"  
16008..16128  
repeat_region    /rpt_family="MLT1F"  
16130..16430  
repeat_region    /rpt_family="AluSp"  
16432..16563  
repeat_region    /rpt_family="(GAA)n"  
16575..16891  
repeat_region    /rpt_family="AluSx"  
16892..16930  
repeat_region    /rpt_family="(GAA)n"  
16989..17073  
repeat_region    /rpt_family="MLT1F"  
17178..17479  
repeat_region    /rpt_family="AluSx"  
17481..18206  
repeat_region    /rpt_family="MER21B"  
18240..18329  
repeat_region    /rpt_family="MLT1F"  
18380..18748  
repeat_region    /rpt_family="MSTD"  
complement(19030..19229)  
repeat_region    /rpt_family="MER20"  
complement(19579..19692)  
misc_feature    /note="predicted exon, program: grail2exons_human_1.3,  
frame: 1, quality: good, score: 53.000"  
repeat_region    complement(19953..20160)  
repeat_region    /rpt_family="MER20"  
complement(20388..20585)  
repeat_region    /rpt_family="MER20"  
20846..20994  
repeat_region    /rpt_family="MR"  
complement(21068..21332)  
repeat_region    /rpt_family="AluSx"  
complement(21372..21671)  
repeat_region    /rpt_family="AluSg"  
complement(21681..21811)  
repeat_region    /rpt_family="FLAM_C"  
21837..22056  
repeat_region    /rpt_family="MER58A"  
22240..22539  
repeat_region    /rpt_family="AluSx"  
22697..22831  
repeat_region    /rpt_family="AluJo"  
22833..23132  
repeat_region    /rpt_family="AluSx"  
23136..23412  
repeat_region    /rpt_family="L1MC3"  
complement(24002..24172)  
repeat_region    /rpt_family="MER20"  
complement(24191..24313)  
repeat_region    /rpt_family="MR"  
24410..24707  
repeat_region    /rpt_family="AluSg"  
complement(24739..24926)  
repeat_region    /rpt_family="MER58A"
```

repeat_region 24928..25229
/rpt_family="AluXx"

Query Match 74.2% Score 23; DB 9; Length 177540;
Best Local Similarity 100.0%; Prid. No. 3.1;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 CTCACATGAGGGTGCAATTGCC 31
|||||
Db 72409 CTCACATGAGGGTGCAATTGCC 72431

RESULT 7
LOCUS AC113495
DEFINITION Mus musculus clone RP23-361011, WORKING DRAFT SEQUENCE, 14 ordered
pieces.
AC113495
AC113495.3 GI:21327431
VERSION
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 194456)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-361011
Unpublished
2 (bases 1 to 194456)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Boguslavskiy,L., Boukhalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Chapel,P., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kasat,A., Karatas,A., Kells,C., Labrecque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Menelus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,T., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomas,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zemock,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 194456)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Chapel,P., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Labrecque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrum,J.,
Menelus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomas,N.,

```

TITLE
JOURNAL

COMMENT
Stojanovic, N., Straus, N., Subramanian, A., Talmas, J., Tesfaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.
Direct Submissions
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gl:2131868.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L23814
Center clone name: 361.0.11
-----
Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 189249 bases at least Q40
Consensus quality: 191283 bases at least Q30
Consensus quality: 192097 bases at least Q20
Insert size: 194000; agarose-bp
Insert size: 193156; sum-of-contigs
Quality coverage: 7.8 in Q20 bases; agarose-fp
Quality coverage: 7.9 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 320: contig of 320 bp in length
* 321 420: gap of 100 bp
* 421 1287: contig of 867 bp in length
* 1288 1387: gap of 100 bp
* 1388 3055: contig of 1668 bp in length
* 3056 3155: gap of 100 bp
* 3156 4819: contig of 1664 bp in length
* 4820 4919: gap of 100 bp
* 4920 7593: contig of 2674 bp in length
* 7594 7693: gap of 100 bp
* 7694 14544: contig of 6851 bp in length
* 14545 14644: gap of 100 bp
* 14645 21443: contig of 6799 bp in length
* 21444 21543: gap of 100 bp
* 21544 34699: contig of 13156 bp in length
* 34700 34799: gap of 100 bp
* 34800 51238: contig of 16439 bp in length
* 51239 51338: gap of 100 bp
* 51339 70476: contig of 19138 bp in length
* 70477 70576: gap of 100 bp
* 70577 123829: contig of 53253 bp in length
* 123830 123929: gap of 100 bp
* 123930 147950: contig of 24021 bp in length
* 147951 148050: gap of 100 bp
* 148051 179165: contig of 31115 bp in length
* 179166 179265: gap of 100 bp
* 179266 194456: contig of 15191 bp in length.
Location/Qualifiers
1. 194456
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-361O11"
/clone_lib="RPCI-23 Female Mouse BAC"

FEATURES
SOURCE
```

```
misc_feature 1. 320
/note="assembly_fragment
clone_end:SP6
vector_side:left"
misc_feature 421.1287
/note="assembly_fragment"
1388.3055
/note="assembly_fragment"
3156.4819
/note="assembly_fragment"
4920.7593
/note="assembly_fragment"
7694.14344
/note="assembly_fragment"
14645.21443
/note="assembly_fragment"
21544.34699
/note="assembly_fragment"
34800.51238
/note="assembly_fragment"
51339.70476
/note="assembly_fragment"
70577.123829
/note="assembly_fragment"
123930.147950
/note="assembly_fragment"
148051.179165
/note="assembly_fragment"
179266.194456
/note="assembly_fragment
clone_end:T7
vector_side:right"
BASE COUNT 52225 a 42789 c 44383 g 53744 t 1315 others
ORIGIN
Query Match 74.2% Score 23; DB 2; Length 194456;
Best Local Similarity 83.9% Pred. No. 3.1;
Matches 26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GCGGAATTCACATGAGGTCATTGGCCC 31
|| ||||| ||||| ||||| |||||
Db 53017 GCTCAATCTCAGATGAGGTGCTAGACCC 53047

RESULT 8
AC026458 153873 bp DNA linear PRI 30-MAR-2002
LOCUS Homo sapiens chromosome 16 clone RP11-10K17, complete sequence.
DEFINITION AC026458
AC026458.7 GI:19848340
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 153873)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 153873)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
AUTHORS 3 (bases 1 to 153873)
TITLE Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 153873)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (30-MAR-2002) Production Sequencing Facility, DOE Joint
```

```
COMMENT
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Mar 30, 2002 this sequence version replaced gi:18767414.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
Source Location/Qualifiers
1. 153873
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-10K17"
BASE COUNT 37256 a 37090 c 39282 g 40245 t
ORIGIN
Query Match 73.5% Score 22.8; DB 9; Length 153873;
Best Local Similarity 92.3% Pred. No. 3.9;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 6 ATTCTCACATGAGGTCATTGGCCC 31
||||| ||||| ||||| |||||
Db 132268 ATTCTCAATGAGGTGATTGGCCC 132293

RESULT 9
AY056037 597 bp mRNA linear PRI 24-JUL-2002
LOCUS Homo sapiens R1g protein mRNA, complete cds.
DEFINITION AY056037
ACCESSION AY056037.1 GI:16555333
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 597)
AUTHORS Ellis,C.A., Vos,M.D., Howell,H., Vallecorsa,T., Fults,D.W. and
Clark,G.J.
TITLE R1g is a novel Ras-related protein and potential neural tumor
suppressor
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (15), 9876-9881 (2002)
AUTHORS 2 (bases 1 to 597)
TITLE Castro,A.F. and Quilliam,L.A.
JOURNAL Identification of R1g, a novel Ras family member sharing homology
with NOY2 and Rap
AUTHORS 3 (bases 1 to 597)
TITLE Direct Submission
JOURNAL Submitted (13-SEP-2001) Biochemistry and Molecular Biology, Indiana
University School of Medicine, 635 Barnhill Drive, MS-4053,
Indianapolis, IN 46202, USA
FEATURES
Source Location/Qualifiers
1. 597
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```

/chrmosome="19"
/map="19p13.3"
/tissue-type="brain; anaplastic oligodendroglioma"
/notes="derived from dbEST AI497811"
1. 597
/notes="Ras family GTPase"
/codon_start=1
/product="Rig protein"
/protein_id="AAL23715.1"
/db_xref="GI:1655334"
/translation="MEQSDNYRVVYFAGAGVGKSSLVLRFKGTFRDYIPIEDTY
ROYISCDKSVCTQITDTTGSHPAPMORLISKGAFILVSVTSKQSLSEGLPIYK
LIVQISGVSDIYVLMVGNKCDREVDREDAOVAQOEKCAFMTSANNVVKEL
FOELLTETFRNNSLINDGRSKOKRTDRVKKCTILM"

BASE COUNT      134 a      178 c      195 g      90 t
ORIGIN
Query Match      71.0%; Score 22; DB 9; Length 597;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      10 TCACATGAGGGTGCAATTTGCC 31
597 TCACATGAGGGTGCAATTTGCC 576

RESULT 10
LOCUS      AY059641      597 bp      mRNA      linear      PRI 29-OCT-2001
DEFINITION      Homo sapiens small GTP-binding tumor suppressor 1 mRNA, complete
cds.
ACCESSION      AY059641
VERSION        AY059641.1 GI:16508175
KEYWORDS
SOURCE
ORGANISM      Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthelia; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 597)
Gong, L. and Wu, K.
Molecular cloning of GBRSL, a novel gene encoding a small
GTP-binding tumor suppressor
Unpublished
2 (bases 1 to 597)
Gong, L.
Direct Submission
Submitted (12-OCT-2001) Dept. of Cardiology, U.T. MD. Anderson
Cancer Center, 1515 Holcombe Blvd-449, Houston, TX 77030, USA
Location/Qualifiers
1. 597
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chrmosome="19"
1. 597
/notes="GBRSL"
/codon_start=1
/product="small GTP-binding tumor suppressor 1"
/protein_id="AAL17968.1"
/db_xref="GI:16508176"
/translation="MEQSDNYRVVYFAGAGVGKSSLVLRFKGTFRDYIPIEDTY
ROYISCDKSVCTQITDTTGSHPAPMORLISKGAFILVSVTSKQSLSEGLPIYK
LIVQISGVSDIYVLMVGNKCDREVDREDAOVAQOEKCAFMTSANNVVKEL
FOELLTETFRNNSLINDGRSKOKRTDRVKKCTILM"

BASE COUNT      134 a      178 c      195 g      90 t
ORIGIN
Query Match      71.0%; Score 22; DB 9; Length 597;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y      10 TCACATGAGGGTGCAATTTGCC 31
597 TCACATGAGGGTGCAATTTGCC 576

```

```

RESULT 11
LOCUS      AC068480
DEFINITION      Homo sapiens chromosome 3 clone RP11-365F21 map 3, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION      AC068480
VERSION        AC068480.1 GI:767964
KEYWORDS
SOURCE
ORGANISM      Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Euthelia; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 67767)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 3, clone RP11-365F21
2 (bases 1 to 67767)
Unpublished
REFERENCE
AUTHORS      Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S.,
Dodgson, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Lacroque, K., Lamazares, R., Landers, J., Lehoczy, S.,
Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meidtm, J., Menus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, D., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, C., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessier, S., Theodore, J., Tjirell, A., Travers, M., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (02-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: U990
Center clone name: 365_F_21
-----
* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1 757: contig of 757 bp in length
* 758 857: gap of 100 bp
* 858 1628: contig of 771 bp in length
* 1629 1728: gap of 100 bp
* 1729 2483: contig of 755 bp in length
* 2484 2583: gap of 100 bp
* 2584 3339: contig of 756 bp in length

```

```
* 3340 3439: gap of 100 bp
* 3440 4197: contig of 758 bp in length
* 4198 4297: gap of 100 bp
* 4298 5035: contig of 738 bp in length
* 5036 5135: gap of 100 bp
* 5136 5864: contig of 729 bp in length
* 5865 5964: gap of 100 bp
* 5965 6700: contig of 736 bp in length
* 6701 6800: gap of 100 bp
* 6801 7544: contig of 744 bp in length
* 7545 7644: gap of 100 bp
* 7645 8404: contig of 760 bp in length
* 8405 8504: gap of 100 bp
* 8505 9238: contig of 734 bp in length
* 9239 9338: gap of 100 bp
* 9339 10092: contig of 754 bp in length
* 10093 10192: gap of 100 bp
* 10193 10956: contig of 764 bp in length
* 10957 11056: gap of 100 bp
* 11057 11809: contig of 753 bp in length
* 11810 11909: gap of 100 bp
* 11910 12658: contig of 749 bp in length
* 12659 12758: gap of 100 bp
* 12759 13505: contig of 747 bp in length
* 13506 13605: gap of 100 bp
* 13606 14343: contig of 738 bp in length
* 14344 14443: gap of 100 bp
* 14444 15177: contig of 734 bp in length
* 15178 15277: gap of 100 bp
* 15278 16020: contig of 743 bp in length
* 16021 16120: gap of 100 bp
* 16121 16874: contig of 754 bp in length
* 16875 16974: gap of 100 bp
* 16975 17719: contig of 745 bp in length
* 17720 17819: gap of 100 bp
* 17820 18558: contig of 739 bp in length
* 18559 18658: gap of 100 bp
* 18659 19414: contig of 756 bp in length
* 19415 19514: gap of 100 bp
* 19515 20260: contig of 746 bp in length
* 20261 20360: gap of 100 bp
* 20361 21109: contig of 749 bp in length
* 21110 21209: gap of 100 bp
* 21210 21959: contig of 750 bp in length
* 21960 22059: gap of 100 bp
* 22060 22808: contig of 749 bp in length
* 22809 22908: gap of 100 bp
* 22909 23660: contig of 752 bp in length
* 23661 23760: gap of 100 bp
* 23761 24517: contig of 757 bp in length
* 24518 24617: gap of 100 bp
* 24618 25378: contig of 761 bp in length
* 25379 25478: gap of 100 bp
* 25479 26227: contig of 749 bp in length
* 26228 26327: gap of 100 bp
* 26328 27083: contig of 756 bp in length
* 27084 27183: gap of 100 bp
* 27184 27929: contig of 746 bp in length
* 27930 28029: gap of 100 bp
* 28030 28778: contig of 749 bp in length
* 28779 28878: gap of 100 bp
* 28879 29642: contig of 764 bp in length
* 29643 29742: gap of 100 bp
* 29743 30497: contig of 755 bp in length
* 30498 30597: gap of 100 bp
* 30598 31347: contig of 750 bp in length
* 31348 31447: gap of 100 bp
* 31448 32194: contig of 747 bp in length
* 32195 32294: gap of 100 bp
* 32295 33047: contig of 753 bp in length
* 33048 33147: gap of 100 bp
* 33148 33879: contig of 732 bp in length
* 33880 33979: gap of 100 bp
```

```
* 33980 34731: contig of 752 bp in length
* 34732 34831: gap of 100 bp
* 34832 35583: contig of 752 bp in length
* 35584 35683: gap of 100 bp
* 35684 36423: contig of 740 bp in length
* 36424 36523: gap of 100 bp
* 36524 37258: contig of 735 bp in length
* 37259 37358: gap of 100 bp
* 37359 38110: contig of 752 bp in length
* 38111 38210: gap of 100 bp
* 38211 38957: contig of 747 bp in length
* 38958 39057: gap of 100 bp
* 39058 39608: contig of 751 bp in length
* 39609 39908: gap of 100 bp
* 39909 40663: contig of 755 bp in length
* 40664 40763: gap of 100 bp
* 40764 41521: contig of 758 bp in length
* 41522 41621: gap of 100 bp
* 41622 42359: contig of 738 bp in length
* 42360 42459: gap of 100 bp
* 42460 43193: contig of 734 bp in length
* 43194 43293: gap of 100 bp
* 43294 44028: contig of 735 bp in length
* 44029 44128: gap of 100 bp
* 44129 44891: contig of 763 bp in length
* 44892 44991: gap of 100 bp
* 44992 45752: contig of 761 bp in length
* 45753 45852: gap of 100 bp
* 45853 46593: contig of 741 bp in length
* 46594 46693: gap of 100 bp
* 46694 47441: contig of 748 bp in length
* 47442 47541: gap of 100 bp
* 47542 48288: contig of 747 bp in length
* 48289 48388: gap of 100 bp
* 48389 49130: contig of 742 bp in length
* 49131 49230: gap of 100 bp
* 49231 49980: contig of 750 bp in length
* 49981 50080: gap of 100 bp
* 50081 50827: contig of 747 bp in length
* 50828 50927: gap of 100 bp
* 50928 51669: contig of 742 bp in length
* 51670 51769: gap of 100 bp
* 51770 52512: contig of 743 bp in length
* 52513 52612: gap of 100 bp
* 52613 53352: contig of 740 bp in length
* 53353 53452: gap of 100 bp
* 53453 54194: contig of 742 bp in length
* 54195 54294: gap of 100 bp
* 54295 55049: contig of 755 bp in length
* 55050 55149: gap of 100 bp
* 55150 55896: contig of 747 bp in length
* 55897 55996: gap of 100 bp
* 55997 56750: contig of 754 bp in length
* 56751 56850: gap of 100 bp
* 56851 57604: contig of 754 bp in length
* 57605 57704: gap of 100 bp
* 57705 58411: contig of 707 bp in length
* 58412 58511: gap of 100 bp
```

Query Match 68.4%; Score 21.2; DB 2; Length 67767;
Best Local Similarity 88.5%; Pred. No. 24;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GCGAATTCACATGAGGATTCATT 26
Db 27478 GCGAATTCACATGAGGAGCGTT 27503

RESULT 12
AC127182/c 185829 bp DNA linear HTG 14-JUL-2002
LOCUS Rattus norvegicus clone CH230-96H21, *** SEQUENCING IN PROGRESS
DEFINITION *** 63 unordered pieces.

ACCESSION AC127182
VERSION AC127182.1 GI:21747570
KEYWORDS HTGS PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 185829)
AUTHORS Mizny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davis,L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,D.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kuesh,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozad,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheswari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okunodu,G.,
Oragunye,N., Oviado,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Soederren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sulton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
TITLE Unpublished
JOURNAL Direct Submission
REFERENCE 2 (bases 1 to 185829)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT ----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: G00S
Center clone name: CH230-96R21
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 131591 bases at least Q40
Consensus quality: 137671 bases at least Q30
Consensus quality: 141923 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 63 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1083: contig of 1083 bp in length
1084 1183: gap of unknown length
1184 2406: contig of 1223 bp in length
2407 2506: gap of unknown length
2507 4126: contig of 1620 bp in length
4127 4226: gap of unknown length
4227 5537: contig of 1311 bp in length
5538 5637: gap of unknown length
5638 7415: contig of 1778 bp in length
7416 7516: gap of unknown length
7516 8626: contig of 1111 bp in length
8627 8726: gap of unknown length
8727 10003: contig of 1277 bp in length
10004 10103: gap of unknown length
10104 11880: contig of 1777 bp in length
11881 11980: gap of unknown length
11981 12985: contig of 1005 bp in length
12986 13085: gap of unknown length
13086 14457: contig of 1372 bp in length
14458 14537: gap of unknown length
14538 15698: contig of 1141 bp in length
15699 15798: gap of unknown length
15799 17182: contig of 1384 bp in length
17183 17282: gap of unknown length
17283 18534: contig of 1252 bp in length
18535 18634: gap of unknown length
18635 19953: contig of 1319 bp in length
19954 20053: gap of unknown length
20054 21317: contig of 1264 bp in length
21318 21417: gap of unknown length
21418 22841: contig of 1424 bp in length
22842 22941: gap of unknown length
22942 24544: contig of 1603 bp in length
24545 24644: gap of unknown length
24645 26066: contig of 1422 bp in length
26067 27351: gap of unknown length
27352 27451: contig of 1185 bp in length
27452 28608: contig of 1157 bp in length
28609 28708: gap of unknown length
28709 30940: contig of 2232 bp in length
30941 31040: gap of unknown length
31041 32702: contig of 1662 bp in length
32703 32802: gap of unknown length
32803 34379: contig of 1577 bp in length
34380 34479: gap of unknown length
34480 36044: contig of 1565 bp in length
36045 36144: gap of unknown length
36145 38735: contig of 2591 bp in length
38736 38835: gap of unknown length
38836 40281: contig of 1446 bp in length
40282 40381: gap of unknown length
40382 42272: contig of 1891 bp in length
42273 42372: gap of unknown length
42373 44841: contig of 2469 bp in length
44842 44941: gap of unknown length
44942 46361: contig of 1420 bp in length
46362 48242: contig of 1781 bp in length
48243 48342: gap of unknown length
48343 49344: contig of 1002 bp in length
49345 52526: gap of unknown length
52526: contig of 3082 bp in length

```

* 52527 52526: gap of unknown length
* 52627 54749: contig of 2123 bp in length
* 54750 54849: gap of unknown length
* 54850 56762: contig of 1913 bp in length
* 56763 56862: gap of unknown length
* 56863 59578: contig of 2716 bp in length
* 59579 59679: gap of unknown length
* 59679 63990: contig of 4311 bp in length
* 63990 64089: gap of unknown length
* 64089 67207: contig of 3118 bp in length
* 67208 70821: gap of unknown length
* 70822 70921: gap of unknown length
* 70921 73810: contig of 2889 bp in length
* 73811 73910: gap of unknown length
* 73911 76613: contig of 2703 bp in length
* 76614 76713: gap of unknown length
* 76714 79312: contig of 2599 bp in length
* 79313 79412: gap of unknown length
* 79413 81823: contig of 2413 bp in length
* 81826 81925: gap of unknown length
* 81926 85896: contig of 3971 bp in length
* 85897 90144: contig of 4148 bp in length
* 90145 90244: gap of unknown length
* 90245 93598: contig of 3354 bp in length
* 93599 93698: gap of unknown length
* 93699 97232: contig of 3534 bp in length
* 97233 97332: gap of unknown length
* 97333 102574: contig of 5242 bp in length
* 102575 105003: contig of 2329 bp in length
* 105004 105103: gap of unknown length
* 105104 109204: contig of 4101 bp in length
* 109205 109304: gap of unknown length
* 109305 114372: contig of 4968 bp in length
* 114373 114372: gap of unknown length
* 114373 118712: contig of 4340 bp in length
* 118713 118812: gap of unknown length
* 118813 123881: contig of 4469 bp in length
* 123882 123881: gap of unknown length
* 123882 127751: contig of 4370 bp in length
* 127752 127851: gap of unknown length
* 127852 133414: contig of 5563 bp in length
* 133415 133514: gap of unknown length
* 133515 137510: contig of 3996 bp in length
* 137511 137610: gap of unknown length
* 137611 142739: contig of 5129 bp in length
* 142740 142839: gap of unknown length
* 142840 147223: contig of 4384 bp in length
* 147224 147323: gap of unknown length

```

Query Match 68.4%; Score 21.2; DB 2; Length 185829;
 Best Local Similarity 88.5%; Pred. No. 26;
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 3 GGAATTCGACATGAGGTCATTG 28
 ||||| ||| ||||| ||||| |||||
 Db 172402 GGAATTCGACATGAGGTCATTG 172377

RESULT 13
 AF348621 3255 bp DNA linear PLN 03-APR-2001
 LOCUS AF348621
 DEFINITION *Emricella unguis* acetamidase (amds) gene, complete cds.
 ACCESSION AF348621
 VERSION AF348621.1 GI:13517938
 KEYWORDS
 SOURCE
 ORGANISM
Emricella unguis.
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiales; Trichocomaceae; *Emricella*.
 REFERENCE
 1 (bases 1 to 3255)

AUTHORS Sharp, J.A., Davis, M.A. and Hynes, M.J.
 TITLE Comparison of amidase genes in Aspergillus species
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 3255)
 AUTHORS Sharp, J.A., Davis, M.A. and Hynes, M.J.
 TITLE Direct Submission
 JOURNAL Submitted (05-FEB-2001) Genetics, University of Melbourne, Royal
 Pde, Parkville, Vic 3010, Australia
 FEATURES
 source
 1..3255
 Location/Qualifiers
 /organism="Emricella unguis"
 /db_xref="taxon:41737"
 <946..>2781
 /gene="amds"
 join(<946..1203,1258..1398,1449..1869,1921..2078,
 2125..>2781)
 /gene="amds"
 /product="acetamidase"
 join(946..1203,1258..1398,1449..1869,1921..2078,
 2125..>2781)
 /gene="amds"
 /codon_start=1
 /product="acetamidase"
 /protein_id="AAK29062.1"
 /db_xref="GI:13517939"

CDS

BASE COUNT 769 a 852 c 820 g 814 t
 ORIGIN
 Query Match 67.7%; Score 21; DB 8; Length 3255;
 Best Local Similarity 82.8%; Pred. No. 24;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 GGAATTCGACATGAGGTCATGCCC 31
 ||||| ||| ||||| ||||| |||||
 Db 1272 GGAATTCGACATGAGGTCATGCCC 1300

RESULT 14
 AP005284 137615 bp DNA linear HTG 31-MAY-2002
 LOCUS AP005284
 DEFINITION *Oryza sativa* (japonica cultivar-group) chromosome 2 clone B1121A12,
 *** SEQUENCING IN PROGRESS ***. In ordered pieces.
 ACCESSION AP005284
 VERSION AP005284.1 GI:21280334
 KEYWORDS HTG; HTGS; PHASE2.
 SOURCE
 Oryza sativa (japonica cultivar-group) (cultivar: Nipponbare) DNA,
 clone: B1121A12.
 Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoae; *Oryza*.
 REFERENCE
 1
 Sasaki, T., Matsumoto, T. and Katayose, Y.
Oryza sativa nipponbare(GAS) genomic DNA, chromosome 2, BAC
 clone: B1121A12
 Published Only in Database (2002)
 2 (bases 1 to 137615)
 Sasaki, T., Matsumoto, T. and Katayose, Y.
 Direct Submission
 Submitted (29-MAY-2002) Takuji Sasaki, National Institute of
 Agricultural Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail: tsasaki@nias.affrc.go.jp, URL: http://irg.p.dna.affrc.go.jp/,

COMMENT

Tel:81-298-38-7441, Fax:81-298-38-7468)
 NOTE: It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have provided by the submitter. This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

source

Location/Qualifiers

1. 137615
 /organism="Oryza sativa (japonica cultivar-group)"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="2"
 /clone="B121A12"

BASE COUNT 37461 a 30979 c 31992 g 37080 t 103 others
 ORIGIN

Query Match 67.7%; Score 21; DB 2; Length 137615;
 Best Local Similarity 82.8%; Pred. No. 32;
 Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 CGGATTCACAGAGGTCATTTGCC 30
 111 1 1111111 111 1111111

DB 23546 CGGACACTCATCTGTCATTTGCC 23574

RESULT 15

AC112349

LOCUS AC112349 155083 bp DNA linear HTG 13-JUL-2002

DEFINITION Rattus norvegicus clone CH230-336K24, *** SEQUENCING IN PROGRESS

AC112349

AC112349.2 GI:21737633
 HTG: HTGS_PHASE1.

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.
 1 (bases 1 to 155083)

REFERENCE
 AUTHORS Wuzny,D.M., Adams,C., Adio-oduola,B., Ali-osman,F.R., Allen,C.,

Alstbrooks,S.L., Amaralunga,H.C., Are,U.R., Ayele,M., Banks,T.,
 Barbieri,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowe,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Chen,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dalhorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gortell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hognes,M., Hollway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kravovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louisleged,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Messy,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Monabdat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenwo,S., Oguh,M., Okwuonu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Severy,G.,
 Scherer,S., Scott,G., Shen,H., Shoochari,N., Sisson,I.,
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasequez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 155083)
 Worley,K.C.
 Direct Submission
 Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 155083)
 Worley,K.C.
 Direct Submission
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 12, 2002 this sequence version replaced gi:18846589.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GPMI
 Center clone name: CH230-336K24
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 115979 bases at least Q40
 Consensus quality: 120179 bases at least Q30
 Consensus quality: 123722 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 54 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1
 1065: contig of 1065 bp in length
 1066: gap of unknown length
 1166: contig of 1519 bp in length
 2684: gap of unknown length
 2685: gap of unknown length
 2785: contig of 1424 bp in length
 4209: gap of unknown length
 4309: contig of 1321 bp in length
 5629: contig of 1749 bp in length
 5730: gap of unknown length
 7478: contig of 1749 bp in length
 7479: gap of unknown length
 7579: contig of 1200 bp in length
 8779: gap of unknown length
 8879: contig of 1529 bp in length
 10408: gap of unknown length
 10508: contig of 1304 bp in length
 11812: gap of unknown length
 11912: contig of 1276 bp in length
 13187: gap of unknown length
 13287: contig of 1195 bp in length
 13288: gap of unknown length
 14482: contig of 1195 bp in length
 14582: gap of unknown length
 14583: contig of 1146 bp in length

15729 15828: gap of unknown length
15829 16849: contig of 1021 bp in length
16850 16949: gap of unknown length
16950 18972: contig of 2023 bp in length
18973 19072: gap of unknown length
19073 20725: contig of 1653 bp in length
20726 20825: gap of unknown length
20826 22038: contig of 1213 bp in length
22039 22138: gap of unknown length
22139 24186: contig of 2048 bp in length
24187 24286: gap of unknown length
24287 25566: contig of 1280 bp in length
25567 25667: gap of unknown length
25668 27089: contig of 1423 bp in length
27090 27189: gap of unknown length
27190 29512: contig of 2323 bp in length
29513 29612: gap of unknown length
29613 32625: contig of 3013 bp in length
32626 32725: gap of unknown length
32726 34261: contig of 1536 bp in length
34262 34361: gap of unknown length
34362 36173: contig of 1812 bp in length
36174 36273: gap of unknown length
36274 38253: contig of 1980 bp in length
38254 38353: gap of unknown length
38354 39956: contig of 1603 bp in length
39957 40056: gap of unknown length
40057 41603: contig of 1547 bp in length
41604 41703: gap of unknown length
41704 43663: contig of 1960 bp in length
43664 43763: gap of unknown length
43764 45801: contig of 2038 bp in length
45802 45901: gap of unknown length
45902 48585: contig of 2684 bp in length
48586 48685: gap of unknown length
48686 50897: contig of 2212 bp in length
50898 53612: contig of 2615 bp in length
53613 53712: gap of unknown length
53713 55684: contig of 1972 bp in length
55685 55784: gap of unknown length
55785 58085: contig of 2301 bp in length
58086 58185: gap of unknown length
58186 61532: contig of 3347 bp in length
61533 61632: gap of unknown length
61633 63874: contig of 2242 bp in length
63875 63974: gap of unknown length
63975 67252: contig of 3278 bp in length
67253 67352: gap of unknown length
67353 69731: contig of 2379 bp in length
69732 69831: gap of unknown length
69832 72178: contig of 2347 bp in length
72179 72278: gap of unknown length
72279 74962: contig of 2684 bp in length
74963 75062: gap of unknown length
75063 78248: contig of 3186 bp in length
78249 78348: gap of unknown length
78349 81862: contig of 3514 bp in length
81863 81962: gap of unknown length
81963 86028: contig of 4066 bp in length
86029 86128: gap of unknown length
86129 88903: contig of 3775 bp in length
88904 90003: gap of unknown length
90004 93625: contig of 3622 bp in length
93626 93725: gap of unknown length
93726 97620: contig of 3895 bp in length
97621 97720: gap of unknown length
97721 103204: contig of 4484 bp in length
103205 103304: gap of unknown length
103305 107646: contig of 5342 bp in length
107647 107746: gap of unknown length
107747 112372: contig of 4626 bp in length
112373 112472: gap of unknown length

112473 117432: contig of 4960 bp in length
117433 117532: gap of unknown length
117533 121402: contig of 3870 bp in length
121403 121502: gap of unknown length
121503 127185: contig of 5683 bp in length
127186 127285: gap of unknown length
127286 132130: contig of 4845 bp in length
132131 132230: gap of unknown length
132231 140048: contig of 7818 bp in length
140049 140148: gap of unknown length
140149 146515: contig of 6367 bp in length
146516 146615: gap of unknown length
146616 155083: contig of 8468 bp in length.

Query Match 66.5%; Score 20.6; DB 2; Length 155083;
Best Local Similarity 85.2%; Pred. No. 52;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 3 GGATTCTCACATGAGGCGTCATTTC 29
DB 4763 GGATTGGGACATGGGGGTGCATTTC 4789

Search completed: March 23, 2003, 16:04:28
Job time : 339.909 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 13:56:37 ; Search time 36.0455 Seconds
(without alignments)
1936.775 Million cell updates/sec

Title: US-09-873-546-3

Perfect score: 31

Sequence: 1 gcggaattccacatgaggtgcatggccc 31

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: N_Geneseq_101002: *
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT: *
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT: *
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT: *
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT: *
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT: *
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT: *
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT: *
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT: *
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT: *
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT: *
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT: *
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT: *
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT: *
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT: *
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT: *
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT: *
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT: *
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT: *
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT: *
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT: *
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT: *
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT: *
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT: *
25: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	74.2	1249	ABO72525	Human MDDT encodin
2	74.2	3061	ABO72648	Human MDDT encodin
3	20	64.5	21521	Human nervous syst
4	20	21521	AAK78175	Human immune/hema
5	19.8	63.9	294	Human OREF polynuc
6	19.8	63.9	433	Human polynucleoti
7	19.8	63.9	527	Human polynucleoti
8	19.8	63.9	710	Human GTP-binding
9	19.8	63.9	2484	Human nervous syst

C 10	19.8	63.9	2484	22	ABA20287	Human nervous syst
C 11	19.4	62.6	3098	20	AAZ00362	Nucleotide sequenc
C 12	19.4	62.6	3098	22	AAF89024	Rat FAP1 coding s
C 13	19.2	61.9	340	24	ABN95508	Gene #2006 used to
C 14	19.2	61.9	340	24	ABK64471	Human benign prost
C 15	19.2	61.9	340	24	ABK64471	Genomic sequence #
C 16	18.6	60.0	24259	22	AAK42245	Tumour suppressor
C 17	18.4	59.4	787	20	AAZ17527	Human gene express
C 18	18.4	59.4	2824	22	AAH18429	Human cDNA sequenc
C 19	18.4	59.4	2936	22	AAH08646	Human cDNA encodin
C 20	18.4	59.4	3098	21	AAA23459	CDNA encoding huma
C 21	18.4	59.4	3844	22	AAH76202	Human drug metabol
C 22	18.4	59.4	3910	22	AAH15815	Human cDNA encodin
C 23	18.4	59.4	3912	22	AAH14598	Human cDNA sequenc
C 24	18.4	59.4	6663	19	AAV27145	Nucleotide sequenc
C 25	18.4	59.4	6663	22	AAD04197	Murine haemopoiet
C 26	18.4	59.4	11832	19	AAV27148	Nucleotide sequenc
C 27	18.4	59.4	11832	22	AAD04198	Murine NR6 genomic
C 28	18.2	58.7	335	22	AAF66850	Novel human polynu
C 29	18.2	58.7	470	24	ABN96653	Gene #3151 used to
C 30	18.2	58.7	470	24	ABK64694	Human benign prost
C 31	18.2	58.7	470	24	ABL68241	Kidney cancer rela
C 32	18.2	58.7	470	24	ABL68479	Kidney cancer rela
C 33	18.2	58.7	470	24	ABL69695	Prostate cancer re
C 34	18.2	58.7	661	24	ABO65887	Arabidopsis thalia
C 35	18.2	58.7	1325	21	AAC32810	Human secreted pro
C 36	18.2	58.7	1809	21	AAV59514	Human immune syste
C 37	18.2	58.7	9817	24	ABL33369	Human brain expres
C 38	18	58.1	353	14	AAO60348	Human RAP11 CDNA
C 39	18	58.1	7653	24	AAO22153	FRAP (tor1) CDNA
C 40	18	58.1	7824	17	AAV38518	EST clone FY127
C 41	18	58.1	7943	19	AAV28518	Human foetal liver
C 42	17.8	57.4	296	20	AAV88071	Human brain expres
C 43	17.8	57.4	514	22	ABA61413	Human bone marrow
C 44	17.8	57.4	514	22	AAK09711	
C 45	17.8	57.4	514	22	AAK35605	

ALIGNMENTS

RESULT 1

ABO72525/c
ID ABO72525 standard; CDNA: 1249 BP.

XX AB072525;

DT 03-SEP-2002 (first entry)

DE Human MDDT encoding cDNA SEQ ID NO 77.

XX Human: MDDT: disease detection and treatment molecule polynucleotide;

XX autolimmune disorder; hepatitis; psoriasis; cancer; AIDS;

XX rheumatoid arthritis; inflammatory; gene therapy; antileukosclerotic;

XX hepatotropic; antineoplastic; antiproliferative; cytoskeletal; anti-HIV;

XX antiallergic; antineoplastic; antineoplastic; antineoplastic; antineoplastic;

XX neuroprotective; antineoplastic; antineoplastic; antineoplastic; antineoplastic;

OS Homo sapiens.

XX WO200240715-A2.

XX 23-MAY-2002.

XX 06-SEP-2001; 2001WO-US27628.

XX 06-SEP-2000; 2000US-230505P.

XX 06-SEP-2000; 2000US-230514P.

XX 06-SEP-2000; 2000US-230515P.

XX 06-SEP-2000; 2000US-230517P.

XX 06-SEP-2000; 2000US-230518P.

XX 06-SEP-2000; 2000US-230519P.

[illegible]

CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal CC and parasitic infections.
CC	Note: The sequence data for this patent did not form part of the CC printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
XX	
SQ	Sequence 21521 BP; 6512 A; 4334 C; 4422 G; 6253 T; 0 other;
<hr/>	
Query Match	64.5%; Score 20; DB 22; Length 21521;
Best Local Similarity	82.1%; Pred.No. 30;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;	
OY	1 GCGGAATTCTCACATGAGGGTCATTGTG 28
Db	16126 GCGGAATTTCCACATGAGGAAAGCAGTTG 16099
<hr/>	
RESULT 4	
AKK78175/C	
ID	AKK78175 standard; DNA; 21521 BP.
XX	
AC	AKK78175;
XX	
DT	07-NOV-2001 (first entry)
XX	
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:32987.
XX	
KW	Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer; cyostatic; gene therapy; vaccine; metastasis; ds.
XX	
OS	Homo sapiens.
XX	
PN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
FF	17-JAN-2001; 2001WO-US01354.
XX	
PR	31-JAN-2000; 2000US-0179065.
PR	04-FEB-2000; 2000US-0180628.
PR	24-FEB-2000; 2000US-0184664.
PR	02-MAR-2000; 2000US-0186350.
PR	16-MAR-2000; 2000US-0189874.
PR	17-MAR-2000; 2000US-0190076.
PR	18-APR-2000; 2000US-0198123.
PR	19-MAY-2000; 2000US-0205515.
PR	07-JUN-2000; 2000US-0209467.
PR	28-JUN-2000; 2000US-0214886.
PR	30-JUN-2000; 2000US-0215135.
PR	07-JUL-2000; 2000US-0216647.
PR	07-JUL-2000; 2000US-0216880.
PR	11-JUL-2000; 2000US-0217487.
PR	11-JUL-2000; 2000US-0217496.
PR	14-JUL-2000; 2000US-0218290.
PR	26-JUL-2000; 2000US-0220963.
PR	26-JUL-2000; 2000US-0220964.
PR	14-AUG-2000; 2000US-0224518.
PR	14-AUG-2000; 2000US-0224519.
PR	14-AUG-2000; 2000US-0225213.
PR	14-AUG-2000; 2000US-0225214.
PR	14-AUG-2000; 2000US-0225215.
PR	14-AUG-2000; 2000US-0225266.
PR	14-AUG-2000; 2000US-0225267.
PR	14-AUG-2000; 2000US-0225268.
PR	14-AUG-2000; 2000US-0225270.
PR	14-AUG-2000; 2000US-0225447.
PR	14-AUG-2000; 2000US-0225757.
PR	14-AUG-2000; 2000US-0225758.
PR	14-AUG-2000; 2000US-0225759.
PR	18-AUG-2000; 2000US-0226279.
PR	22-AUG-2000; 2000US-0226681.
PR	22-AUG-2000; 2000US-0226868.
PR	22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234297.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235184.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246539.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI; 2001-483426/52.
 Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 useful for preventing, diagnosing and/or treating cancers and
 metastasis -
 XX
 PS Disclosure; SEQ ID NO 32987; 3071pp + Sequence listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patients own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 21521 BP; 6512 A; 4334 C; 4422 G; 6253 T; 0 other;

Query Match 64.5%; Score 20; DB 22; Length 21521;
 Best Local Similarity 82.1%; Pred No. 30;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GCGGAATTCACATGAGGCGTCATTG 28
 Db 16126 GCGGAATTCACATGAGGAGCAGTTG 16099

RESULT 5
 ABN20378/c

ID ABN20378 standard; cDNA: 294 BP.
XX
AC ABN20378;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX polynucleotide sequence SEQ ID NO:9233.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis; gene; ss.
XX
OS Homo sapiens.
XX
WO200192523-A2.
XX
06-DEC-2001.
XX
29-MAY-2001; 2001WO-US10836.
XX
30-MAY-2000; 2000US-206132P.
XX
29-AUG-2000; 2000US-228716P.
XX
(CURA-) CURAGEN CORP.
XX
Shinkets RA, Leach MD:
XX
WPI; 2002-106308/14.
XX
P-PSDB; ABP04626.
XX
Novel human polypeptides and polynucleotides useful for diagnosing,
XX
preventing and treating cardiovascular disease, neurodegenerative,
XX
hyperproliferative disorders and autoimmune disorders -
XX
Disclosure: SEQ ID 9233; 1037pp; English.
XX
PS The present invention describes substantially purified human proteins
XX (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
XX in the specification). ABN15762 to ABN27252 encode the human ORFX
XX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
XX treating or preventing a pathology associated with an ORFX-associated
XX disorder in humans, and in the manufacture of a medicament for treating a
XX syndrome associated with ORFX-associated disorder. ORFX polynucleotide
XX sequences can be used in gene therapy. ORFX sequences can be used in the
XX treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
XX psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
XX osteoarthritis, neurodegenerative disorders, disorders related to organ
XX transplantation, cardiovascular diseases, diabetes mellitus, systemic
XX lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
XX storage disease, various immune deficiencies and disorders, infectious
XX diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
XX arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
XX disease and autoimmune inflammatory eye disease. ORFX proteins are also
XX useful for treating burns, incisions, ulcers, for treating osteoporosis,
XX bone degenerative disorders, or periodontal disease, and for gut
XX protection or regeneration and treatment of lung or liver fibrosis,
XX reperfusion injury in various tissues and conditions resulting from
XX systemic cytokine damage.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIRO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 294 BP; 59 A; 83 C; 84 G; 67 T; 1 other;
Query Match 63.9%; Score 19.8; DB 24; Length 294;
Best Local Similarity 77.4%; Pred. No. 16;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 GCGGAATTCACATGAGGTCATTGGCC 31
II IIIII IIII I IIII IIIII I
DB 225 GCCGAATTCACACAGGTGCGTCATTGGCC 195
RESULT 6
ID AAK52951/c
XX AAK52951 standard; cDNA: 433 BP.
XX
AC AAK52951;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 2480.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
WO200157190-A2.
XX
09-AUG-2001.
XX
05-FEB-2001; 2001WO-US04098.
XX
03-FEB-2000; 2000US-0496914.
XX
27-APR-2000; 2000US-0560875.
XX
20-JUN-2000; 2000US-0598075.
XX
19-JUL-2000; 2000US-0620325.
XX
01-SEP-2000; 2000US-0654936.
XX
15-SEP-2000; 2000US-0663561.
XX
20-OCT-2000; 2000US-0693325.
XX
30-NOV-2000; 2000US-0728422.
XX
(HYSE-) HYSEQ INC.
XX
Tang YT, Liu C, Dermanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX
Pi Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX
Pi Xue AJ, Yang Y, Wejthman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
XX
P-PSDB; AAM79818.
XX
Nucleic acids encoding polypeptides with cytokine-like activities,
XX
useful in diagnosis and gene therapy -
XX
Claim 1; Page 4758-4759; 6221pp; English.
XX
PS The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activity/inhibit activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
XX
SQ Sequence 433 BP; 90 A; 145 C; 113 G; 84 T; 1 other;
Query Match 63.9%; Score 19.8; DB 22; Length 433;
Best Local Similarity 91.3%; Pred. No. 17;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 9 CTCACATGAGGTCGATTGGCC 31

Db 348 CTCACAGAGTGTGCATTGGCC 326

||||| || |||||||||

RESULT 7

AAK51967/c
ID AAK51967 standard; cDNA; 527 BP.

XX AAK51967;

DT 06-NOV-2001 (first entry)

DE Human polynucleotide SEQ ID NO 512.

XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.

PN WO200157190-A2.

PD 09-AUG-2001.

PF 05-FEB-2001; 2001WO-US04098.

PR 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663325.

PR 20-OCT-2000; 2000US-0693325.

PR 30-NOV-2000; 2000US-0728422.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao OA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejthman T, Goodrich R;

DR WPI: 2001-476283/51.
P-PSDB; AAM78834.

PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy

XX Claim 1; Page 1850-1851; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78333-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.

XX Sequence 527 BP; 111 A; 159 C; 129 G; 128 T; 0 other;

Query Match 63.9%; Score 19.8; DB 22; Length 527;

Best Local Similarity 91.3%; Pred. No. 18;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

9 CTCACATGAGGTCGATTGGCC 31

||||| || |||||||||

Db 347 CTCACAGAGTGTGCATTGGCC 325

RESULT 8

AAF58344/c
ID AAF58344 standard; cDNA; 710 BP.

XX AAF58344;

DT 19-APR-2001 (first entry)

DE Human GTP-binding associated protein #44 coding sequence.

XX Human: guanosine triphosphate binding associated protein; GTP; GBAP;
KM inflammation; AIDS; Addison's disease; anaemia; arteriosclerosis; asthma;
KW autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;
KW osteoporosis; psoriasis; ss.

OS Homo sapiens.

PN WO200105970-A2.

PD 25-JAN-2001.

PF 19-JUL-2000; 2000WO-US19698.

PR 19-JUL-1999; 99US-0144595.

PR 23-AUG-1999; 99US-0150460.

PR 15-OCT-1999; 99US-0159849.

PA (INCYTE-) INCYTE GENOMICS INC.

PI Yue H, Tang YT, Bandman O, Hillman JL, Lai P, Au-Young J;
PI Reddy R, Yang J, Baughn MR, Lu DM, Azimzai Y, Patterson C;

DR WPI: 2001-091972/10.
P-PSDB; AAB68544.

PT New guanosine triphosphate-binding associated proteins (GBAP) and their
PT encoding nucleic acids, useful for treating and/or diagnosing diseases
PT associated with GBAP expression, such as cancer, diabetes and asthma

XX Claim 5; Page 216; 233pp; English.

XX The present invention relates to novel human guanosine triphosphate
CC (GTP)-binding associated proteins (GBAPs; AAB68501-AAB68566) and their
CC coding sequences (AAF58301-AAF58366). The proteins and coding sequences
CC of the present invention are useful for treating a variety of disorders
CC including inflammation, AIDS, Addison's disease, anaemia,
CC arteriosclerosis, asthma, autoimmune disorders, Grave's disease,
CC hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and
CC psoriasis.

XX Sequence 710 BP; 145 A; 290 C; 153 G; 122 T; 0 other;

Query Match 63.9%; Score 19.8; DB 22; Length 710;

Best Local Similarity 91.3%; Pred. No. 19;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

9 CTCACATGAGGTCGATTGGCC 31

||||| || |||||||||

Db 290 CTCACAGAGTGTGCATTGGCC 268

RESULT 9

ABA20286/c
ID ABA20286 standard; DNA; 2484 BP.

XX ABA20286;

DT 23-JAN-2002 (first entry)

DE Human nervous system related polynucleotide SEQ ID NO 12617.

PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-541565/60.
XX
PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
PS Disclosure; SEQ ID NO 12617; 1701bp + Sequence Listing; English.
XX
CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
CC (ABR14678-ABR18001) useful for preventing, treating or ameliorating
CC medical conditions e.g., by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2484 BP; 440 A; 906 C; 734 G; 404 T; 0 other;

Query Match 63.9%; Score 19.8; DB 22; Length 2484;
Best Local Similarity 91.3%; Pred. No. 24;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 9 CTCACATGAGGTCATTGCC 31
Db 2036 CTCACAGGATGTCATTGCC 2014
||||| |||||||

RESULT 10
ABA20287/c
ID ABA20287 standard; DNA: 2484 BP.
XX
AC ABA20287;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 12618.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antitickling; antianaemic; antiarthritic; cancer;
KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antileucic; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.

XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225216.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231415.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.

```
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
XX

PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
XX
XX
XX Disclosure; SEQ ID NO 12618; 1701pp + Sequence Listing; English.
XX
XX
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins
XX (ABA14678-ABA18001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemia;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pt_sequences.
XX
XX
XX Sequence 2484 BP; 440 A; 906 C; 734 G; 404 T; 0 other;
XX
XX
XX Query Match 63.9%; Score 19.8; DB 22; Length 2484;
XX Best Local Similarity 91.3%; Pred. No. 24;
XX Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX
XX 9 CTCACATGAGGTCGATTGCCC 31
XX
XX
XX Db 2036 CTCACAGATGTCGATTGCC 2014
XX
XX
XX
XX RESULT 11
XX AAZ00362/C
XX ID AAZ00362 standard; DNA; 3098 BP.
XX
XX
XX AC AAZ00362;
XX
XX
XX DT 26-OCT-1999 (first entry)
XX
XX
XX DE Nucleotide sequence of rat rnfATP1.
XX
XX
XX KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA;
XX fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
XX
XX
XX OS Rattus norvegicus.
XX
XX
XX PN WO9936537-A2.
XX
XX
XX PD 22-JUL-1999.
XX
XX
XX PF 14-JAN-1999; 99WO-US00182.
XX
XX
XX PR 14-JAN-1999; 99US-0232201.
XX PR 15-JAN-1998; 98US-0071374.
XX PR 20-JUL-1998; 98US-0093491.
XX PR 04-DEC-1998; 98US-0110941.
XX PR 14-JAN-1999; 99US-0232195.
XX PR 14-JAN-1999; 99US-0232197.
XX PR 14-JAN-1999; 99US-0232200.
XX
XX
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
XX
XX
XX PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
XX
XX WPI; 1999-444398/37.
XX
XX
XX P-PSDB; AA14952.
XX
```


CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
CC XX

SO Sequence 340 BP; 84 A; 79 C; 98 G; 76 T; 3 other;

Query Match 61.9%; Score 19.2; DB 24; Length 340;
Best Local Similarity 87.5%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 7 TTCTCATGAGGTCGTCATTGCC 30
||| || ||||| ||||| |||||
Db 238 TTCTTCAGAGGTCGTCATTGCC 261

RESULT 14
ABK64471
ABK64471 standard; DNA; 340 BP.

AC ABK64471;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human benign prostatic hyperplasia gene #366.
XX
KW Human; benign prostatic hyperplasia; BPH; prostate cancer; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200212440-A2.
XX
PD 14-FEB-2002.
XX
PF 07-AUG-2001; 2001WO-US24708.
XX
PR 07-AUG-2000; 2000US-223323P.
XX
PR 05-JUN-2001; 2001US-0873319.
XX
PA (GENE-) GENE LOGIC INC.
XX (NISB) JAPAN TOBACCO INC.
XX
PI Munger WE, Kulkarni P, Getzenberg RH, Waga I, Yamamoto J;
XX
DR WPI; 2002-257476/30.
XX
XX Identifying drugs for and diagnosing benign prostatic hyperplasia, by
XX detecting expression levels of one or more genes in prostate cells from
XX patient that are differentially regulated compared to normal prostate
XX cells
XX
PS Disclosure; Page 207; 444pp; English.
XX
XX The invention relates to a method of diagnosing (I) the onset or
XX progression of benign prostatic hyperplasia (BPH), or screening (II) for
XX or identifying an agent that modulates the onset or progression of BPH.
XX The method is based on changes in gene expression in BPH tissue isolated
XX from patients exhibiting different clinical states of prostate
XX hyperplasia as compared to normal prostate tissue. (I) comprises
XX detecting the expression levels of one or more genes in prostate cells
XX from the subject that are differentially regulated compared to normal
XX prostate cells. (II) comprises preparing a first gene expression profile
XX of BPH cells or BPH-like cell population, exposing the cells to the
XX agent, preparing a second gene expression profile of the agent exposed
XX cells, and comparing the first and second gene expression profiles
XX (I) is useful for diagnosing the onset or progression of BPH. (II) is
XX useful for identifying an agent that modulates the onset or progression
XX of BPH. The methods are useful to present information identifying
XX the expression level in a tissue or cells, by comparing the expression
XX level of genes given in the specification in the tissue or cells to the
XX level of expression of gene in the database, and displaying the

CC expression levels of at least one gene in the tissue or cell sample
CC compared to the expression level in BPH. Agents using (II) are useful for
CC treating BPH or prostate cancer. ABK64106-ABK64860 represent human
CC benign prostatic hyperplasia gene sequences of the invention.
CC XX

SO Sequence 340 BP; 84 A; 79 C; 98 G; 76 T; 3 other;

Query Match 61.9%; Score 19.2; DB 24; Length 340;
Best Local Similarity 87.5%; Pred. No. 31;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 7 TTCTCATGAGGTCGTCATTGCC 30
||| || ||||| ||||| |||||
Db 238 TTCTTCAGAGGTCGTCATTGCC 261

RESULT 15
ABK42245/C
ID ABK42245 standard; DNA; 12932 BP.

AC ABK42245;
XX
DT 21-MAY-2002 (first entry)
XX
DE Genomic sequence #144 encoding novel human connective tissue polypeptide.
XX
KW Human; connective tissue related disorder; cancer; gene therapy;
XX
KW cytostatic; gene; ds.
XX
OS Homo sapiens.
XX
PN WO20015343-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01322.
XX
PR 31-JAN-2000; 2000US-0179065.
XX
PR 04-FEB-2000; 2000US-0180628.
XX
PR 24-FEB-2000; 2000US-0184664.
XX
PR 02-MAR-2000; 2000US-0186350.
XX
PR 16-MAR-2000; 2000US-0189874.
XX
PR 17-MAR-2000; 2000US-0190076.
XX
PR 18-APR-2000; 2000US-0198123.
XX
PR 19-MAY-2000; 2000US-0205515.
XX
PR 07-JUN-2000; 2000US-0209467.
XX
PR 28-JUN-2000; 2000US-0214886.
XX
PR 30-JUN-2000; 2000US-0215135.
XX
PR 07-JUL-2000; 2000US-0216647.
XX
PR 11-JUL-2000; 2000US-0216880.
XX
PR 11-JUL-2000; 2000US-0217487.
XX
PR 11-JUL-2000; 2000US-0217496.
XX
PR 14-JUL-2000; 2000US-0218290.
XX
PR 26-JUL-2000; 2000US-0220963.
XX
PR 26-JUL-2000; 2000US-0220964.
XX
PR 14-AUG-2000; 2000US-0224518.
XX
PR 14-AUG-2000; 2000US-0224519.
XX
PR 14-AUG-2000; 2000US-0225213.
XX
PR 14-AUG-2000; 2000US-0225214.
XX
PR 14-AUG-2000; 2000US-0225266.
XX
PR 14-AUG-2000; 2000US-0225267.
XX
PR 14-AUG-2000; 2000US-0225268.
XX
PR 14-AUG-2000; 2000US-0225270.
XX
PR 14-AUG-2000; 2000US-0225447.
XX
PR 14-AUG-2000; 2000US-0225757.
XX
PR 14-AUG-2000; 2000US-0225758.
XX
PR 14-AUG-2000; 2000US-0225759.
XX
PR 18-AUG-2000; 2000US-0226279.
XX
PR 22-AUG-2000; 2000US-0226681.
XX
PR 22-AUG-2000; 2000US-0226681.
XX
PR 22-AUG-2000; 2000US-0227182.
XX
PR 23-AUG-2000; 2000US-0227009.
XX
PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246533.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX
DR WPI; 2001-565190/63.
XX
XX
PT Nucleic acid encoding novel connective tissue associated polypeptides,
PT used in diagnosing, preventing, treating or ameliorating a disorder
PT such as cancer or rheumatoid arthritis -
XX
XX
PS Disclosure: SEQ ID No 1132; 673pp; English.
XX
XX
CC The present invention relates to the isolation of novel human connective
CC tissue related polypeptides (AA086435-AA086923) and the polynucleotide
CC (cDNA and genomic) sequences encoding them. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of diseases associated with connective tissue(s), including
CC cancer. The polynucleotide sequences of the invention are also useful
CC in gene therapy. ABR42102-ABR43116 represent genomic sequences encoding
CC the novel human connective tissue related polypeptides.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 12932 BP; 4081 A; 2677 C; 2632 G; 3542 T; 0 other;

Search completed: March 23, 2003, 15:20:04
Job time : 49.5455 secs

Best Match 61.3%; Score 19; DB 23; Length 12932;
Query Locality 81.5%; Pred. No. 81;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
OY 3 GGAATTCACATGAGGTCATTGC 29
DB 3205 GGGATTCACATGACAGTCATTGC 3179
II IIII IIIII IIIII IIIII

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 15:14:42 ; Search time 8.18182 Seconds
(Without alignments)
2938.192 Million cell updates/sec

Title: US-09-873-546-3
Perfect score: 31
Sequence: 1 gcggaattccacatgaggtgcatgtcc 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 538826 seqs, 387737923 residues

Total number of hits satisfying chosen parameters: 1077652

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	19.2	61.9	340	10	US-09-880-107-2006
2	19	61.3	12932	10	US-09-764-847-1132
3	18.4	59.4	2936	10	US-09-747-835A-112
4	18.4	59.4	6663	10	US-09-037-657-28
5	18.4	59.4	11832	10	US-09-037-657-38
6	18.2	58.7	470	10	US-09-969-708-561
7	18.2	58.7	470	10	US-09-880-107-3150
8	18.2	58.7	661	10	US-09-770-149-464
9	18	58.1	7653	9	US-09-950-634-1
10	17.8	57.4	292	10	US-09-783-590-2469
11	17.8	57.4	296	9	US-10-040-739-549
12	17.8	57.4	514	10	US-09-864-761-12738
13	17.8	57.4	3666	9	US-09-895-913A-245
14	17.8	57.4	31718	9	US-09-764-872-812
15	17.8	57.4	31718	9	US-09-764-872-813
16	17.4	56.1	269	10	US-09-878-574-9123
17	17.4	56.1	485	10	US-09-815-242-2683
18	17.4	56.1	705	10	US-09-815-242-8867
19	17.4	56.1	705	10	US-09-815-242-8941

20	17.4	56.1	1539	9	US-09-938-842A-176	Sequence 176, App
21	17.4	56.1	32193	10	US-09-764-877-2623	Sequence 2623, Ap
22	17.2	55.5	466	10	US-09-864-761-5439	Sequence 5439, Ap
23	17.2	55.5	716	10	US-09-070-927A-866	Sequence 866, App
24	17.2	55.5	1345	10	US-09-764-864-315	Sequence 315, App
25	17.2	55.5	1841	10	US-09-764-864-735	Sequence 735, App
26	17.2	55.5	3633	12	US-10-044-090-123	Sequence 123, App
27	17.2	55.5	6633	10	US-09-995-542-1	Sequence 1, Appli
28	17.2	55.5	7621	10	US-09-954-456-307	Sequence 307, App
29	17.2	55.5	7621	10	US-09-954-456-2209	Sequence 2209, Ap
30	17	54.8	390	10	US-09-949-889-2	Sequence 2, Appli
31	17	54.8	1017	9	US-09-938-842A-60	Sequence 60, Appl
32	17	54.8	1710	9	US-09-912-628-2	Sequence 2, Appli
33	17	54.8	1743	9	US-10-006-856A-235	Sequence 235, App
34	17	54.8	1743	9	US-10-184-644-451	Sequence 451, App
35	17	54.8	7838	10	US-09-761-466-4	Sequence 4, Appli
36	17	54.8	148567	9	US-10-254-869-3	Sequence 3, Appli
37	17	54.8	148567	10	US-09-801-876B-3	Sequence 3, Appli
38	16.8	54.2	286	9	US-10-040-739-75	Sequence 75, Appl
39	16.8	54.2	362	10	US-09-878-574-1089	Sequence 1089, Ap
40	16.8	54.2	365	9	US-09-756-854-6	Sequence 6, Appli
41	16.8	54.2	365	9	US-10-041-574-6	Sequence 6, Appli
42	16.8	54.2	400	10	US-09-864-761-10344	Sequence 10344, A
43	16.8	54.2	410	10	US-09-967-768A-88	Sequence 88, Appl
44	16.8	54.2	420	10	US-09-864-761-23266	Sequence 23266, A
45	16.8	54.2	444	10	US-09-960-352-5693	Sequence 5693, Ap

ALIGNMENTS

RESULT 1
US-09-880-107-2006
Sequence 2006, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-MO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2006
LENGTH: 340
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 H81070
NAME/KEY: unsure
LOCATION: (1) (340)
OTHER INFORMATION: n = a or c or g or t
US-09-880-107-2006
Query Match 61.9%; Score 19.2; DB 10; Length 340;
Best Local Similarity 87.5%; Pred. No. 13;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 7 TTCTCAGATGAGGTCATTTGCC 30
Db 238 TTCTTCAGAGGTCATTTGCC 261
RESULT 2
US-09-764-847-1132/c
Sequence 1132, Application US/09764847

```
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; PRIOR APPLICATION DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1132
; LENGTH: 12932
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1132

Query Match
Best Local Similarity 61.3%; Score 19; DB 10; Length 12932;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

3 GGAATTCTCAGATGAGGTCGATTGGC 29
|| |||| |||| |||| |||| ||||
Db 3205 GGGATTCTGACATGACAGTCGATTGGC 3179

RESULT 3
US-09-747-835A-12/c
; Sequence 12, Application US/09747835A
; Patent No. US20020146692A1
; GENERAL INFORMATION:
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Taniguchi, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Wang, Dunru
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyun
; APPLICANT: Asundi, Vinod
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE
; FILE REFERENCE: HYS-37CIP
; CURRENT APPLICATION NUMBER: US/09/747,835A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/729,739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653,450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620,312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 12
; LENGTH: 2936
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (135)..(1601)
US-09-747-835A-12

Query Match
Best Local Similarity 59.4%; Score 18.4; DB 10; Length 2936;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

2 CGGAATTCACATGAGGTCGATTGGC 29
|| || |||| |||| |||| |||| ||||
```

```
Db 2342 CGTCACTCTCAGAGGAGATTCATTGGC 2315

RESULT 4
US-09-037-657-28/c
; Sequence 28, Application US/09037657A
; Patent No. US20020045741A1
; GENERAL INFORMATION:
; APPLICANT: Hilton, Douglas J.
; APPLICANT: Nicola, Nicos A.
; APPLICANT: Farley, Allison
; APPLICANT: Wilson, Tracy
; APPLICANT: Zhang, Jian-Guo
; APPLICANT: Alexander, Warren
; APPLICANT: Rakar, Steven
; APPLICANT: Fabril, Louis
; APPLICANT: Kojima, Tetsuo
; APPLICANT: Maeda, Masatsugu
; APPLICANT: Kikuchi, Yasufumi
; APPLICANT: Nash, Andrew
; TITLE OF INVENTION: A NOVEL HAEMPOIETIN RECEPTOR AND GENETIC SEQUENCES
; FILE REFERENCE: DAVIES COLLISON CAVE (CIP)
; CURRENT APPLICATION NUMBER: US/09/037,657A
; EARLIER FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 08/928,720
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 28
; LENGTH: 6663
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: Murine NR6 gene
US-09-037-657-28

Query Match
Best Local Similarity 59.4%; Score 18.4; DB 10; Length 6663;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

3 GGAATTCACATGAGGTCGATTGGC 30
|| |||| |||| || || || || || ||
Db 594 GGAATTCACATGAGTCGCTTGGC 567

RESULT 5
US-09-037-657-38/c
; Sequence 38, Application US/09037657A
; Patent No. US20020045741A1
; GENERAL INFORMATION:
; APPLICANT: Hilton, Douglas J.
; APPLICANT: Nicola, Nicos A.
; APPLICANT: Farley, Allison
; APPLICANT: Wilson, Tracy
; APPLICANT: Zhang, Jian-Guo
; APPLICANT: Alexander, Warren
; APPLICANT: Rakar, Steven
; APPLICANT: Fabril, Louis
; APPLICANT: Kojima, Tetsuo
; APPLICANT: Maeda, Masatsugu
; APPLICANT: Kikuchi, Yasufumi
; APPLICANT: Nash, Andrew
; TITLE OF INVENTION: A NOVEL HAEMPOIETIN RECEPTOR AND GENETIC SEQUENCES
; FILE REFERENCE: DAVIES COLLISON CAVE (CIP)
; CURRENT APPLICATION NUMBER: US/09/037,657A
; EARLIER FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 08/928,720
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 38
```



```

1      Falli, Amedeo F.
2      Cagliano, Thomas J.
3      Nakanishi, Koji
4      Chen, Yangtzu
5      TITLE OF INVENTION: EFFECTOR PROTEINS OF RAPAMYCIN
6      NUMBER OF SEQUENCES: 23
7      CORRESPONDENCE ADDRESS:
8      ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
9      Dunnet, L.L.P.
10     STREET: 1300 I Street, N.W.
11     CITY: Washington
12     STATE: DC
13     COUNTRY: USA
14     ZIP: 20005-3315
15     COMPUTER READABLE FORM:
16     MEDIUM TYPE: floppy disk
17     COMPUTER: IBM PC compatible
18     OPERATING SYSTEM: PC-DOS/MS-DOS
19     SOFTWARE: PatentIn Release #1.0, Version #1.30
20     CURRENT APPLICATION DATA:
21     APPLICATION NUMBER: US/09/950,654
22     FILING DATE: 13-Sep-2001
23     CLASSIFICATION: unknown>
24     PRIOR APPLICATION DATA:
25     APPLICATION NUMBER: US 08/471,112
26     FILING DATE: 06-JUN-1995
27     APPLICATION NUMBER: US 08/384,524
28     FILING DATE: 13-FEB-1995
29     APPLICATION NUMBER: US 08/312,023
30     FILING DATE: 26-SEP-1995
31     APPLICATION NUMBER: US 08/207,975
32     FILING DATE: 08-MAR-1994
33     ATTORNEY/AGENT INFORMATION:
34     NAME: Slekman, Michael T.
35     REGISTRATION NUMBER: 56,276
36     REFERENCE/DOCKET NUMBER: 01142.0058-00000
37     TELECOMMUNICATION INFORMATION:
38     TELEPHONE: 202-408-4000
39     TELEFAX: 202-408-4400
40     INFORMATION FOR SEQ ID NO: 1:
41     SEQUENCE CHARACTERISTICS:
42     LENGTH: 7653 base pairs
43     TYPE: nucleic acid
44     STRANDEDNESS: double
45     TOPOLOGY: linear
46     MOLECULE TYPE: cDNA to mRNA
47     SEQUENCE DESCRIPTION: SEQ ID NO: 1:
48     US-09-950-634-1
49     Query Match          58.1%; Score 18; DB 9; Length 7653;
50     Best Local Similarity 80.8%; Pred. No. 80;
51     Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
52     QY      2 CGAATTCTCACATGAGGTGCATT 27
53           | ||||| ||||| ||||| |||||
54     Db      3128 CAGAATTCCTCATGAGGTGACTAT 3103
55
56 RESULT 10
57 US-09-783-590-2469/c
58 : Sequence 2469, Application US/09783590
59 : Patent No. US20020110850A1
60 : GENERAL INFORMATION:
61 : APPLICANT: Dillon, Patrick J.
62 : APPLICANT: Haseltine, William A.
63 : APPLICANT: Li, Haodong
64 : APPLICANT: Rosen, Craig A.
65 : APPLICANT: Ruden, Steven M.
66 : TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
67 : FILE REFERENCE: PD-16.2C1
68 : CURRENT APPLICATION NUMBER: US/09/783,590
69 : CURRENT FILING DATE: 2000-02-15
70 : PRIOR APPLICATION NUMBER: 08/420,856

```

```

PRIOR FILING DATE: 1995-04-12
PRIOR APPLICATION NUMBER: 08/346,731
PRIOR FILING DATE: 1994-11-21
NUMBER OF SEQ ID NOS: 12465
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 2469
LENGTH: 292
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (10)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (59)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (153)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (179)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (220)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (226)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (246)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-2469

Query Match          57.4%; Score 17.6; DB 10; Length 292;
Best Local Similarity 90.5%; Pred. No. 57;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps

OY      11 CACATGAGGTGCATTGGCC 31
        ||||| || ||||| ||||| |||||
DB      287 CACAGCATGTGCATTGGCC 267

RESULT 11
US-10-040-739-549
; Sequence 549, Application US/10040739
; Patent No. US20020173635A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John
Lavallie, Edward
Racie, Lisa
Merberg, David
Trecey, Maurice
Spaulding, Vikki
TITLE OF INVENTION: SECRETED, EXPRESSED SEQUENCE TAGS
NUMBER OF SEQUENCES: 1519
CORRESPONDENCE ADDRESS:
ADDRESSER: Genetics Institute, Inc.
STREET: 87 CambridgePark Drive
CITY: Cambridge
STATE: Massachusetts
COUNTRY: U.S.A
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/040,739
FILING DATE: 07-Jan-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:

```

RESULT 12
US-09-864-761-12738
Sequence 12738, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEOTIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT FILING DATE: 2001-05-23
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21

```

RESULT 13
US-09-895-913A-245/c
: Sequence 245, Application US/09895913A
: Patent No. US20020160456A1
: GENERAL INFORMATION:
: APPLICANT: Kleantous, Harold
: APPLICANT: Al-Garawi, Amal
: APPLICANT: Miller, Charles
: APPLICANT: Tomb, Jean Francois
: APPLICANT: Oomen, Raymond P.
: TITLE OF INVENTION: Identification of Polynucleotides
: TITLE OF INVENTION: Encoding No. US20020160456a1 Helicobacter Polypeptides in t
: TITLE OF INVENTION: Genome
: FILE REFERENCE: 06132/043002
: CURRENT APPLICATION NUMBER: US/09/895,913A
: PRIORITY FILING DATE: 2001-06-29
: PRIOR APPLICATION NUMBER: US 08/861,227
: PRIOR FILING DATE: 1997-06-24
: NUMBER OF SEQ ID NOS: 368
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 245
: LENGTH: 3666
: TYPE: DNA
: ORGANISM: Helicobacter pylori
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (54)...(3608)
US-09-895-913A-245

Query Match          57.4%; Score 17.8; DB 9; Length 3666;
Best Local Similarity 75.9%; Pred. No. 88;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

xy      3  GGAATTCACATGAGGGTCATTGGCC 31
      11 11111111111111111111
Db      364 GGCACGCTCATATCAGGAGCATTTTCCC 336

RESULT 14
US-09-764-872-812
: Sequence 812, Application US/09764872
: Publication No. US20030050231A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: APPLICANT: Nucleic Acids, Proteins, and Antibodies
: TITLE OF INVENTION:

```

FILE REFERENCE: PA125
CURRENT APPLICATION NUMBER: US/09/764,872
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 957
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 812
LENGTH: 31718
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-872-812

Query Match 57.4%; Score 17.8; DB 9; Length 31718;
Best Local Similarity 75.9%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 GCGGAATTCACATGAGGGTGCAATTGC 29
DB 24277 GAGGACTTCCACAGAGAGGTGGCATCTGC 24305

RESULT 15

US-09-764-872-813
Sequence 813, Application US/09/764872
Publication No. US20030050231A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA125
CURRENT APPLICATION NUMBER: US/09/764,872
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 957
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 813
LENGTH: 31718
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-872-813

Query Match 57.4%; Score 17.8; DB 9; Length 31718;
Best Local Similarity 75.9%; Pred. No. 1.3e+02;
Matches 22; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 GCGGAATTCACATGAGGGTGCAATTGC 29
DB 24277 GAGGACTTCCACAGAGAGGTGGCATCTGC 24305

Search completed: March 23, 2003, 17:17:01
Job time : 29.1818 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 23, 2003, 14:20:51 ; Search time 201.545 Seconds
(without alignments)
2491.051 Million cell updates/sec

Title: US-09-873-546-3
Perfect score: 31
Sequence: 1 gcgcattccatccatgcaggtcattgcgcc 31

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estcov:*
5: em_estov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_huv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_trod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	23	74.2	420	AJ474604
2	23	74.2	574	AV835621
3	23	74.2	1119	BE885242
4	21.2	68.4	757	AQ349812
5	21	67.7	984	BF796120
6	20.6	66.5	385	BG955224

Result No.	Score	Query Match Length	ID	Description
7	20.6	66.5	528	AM128188
8	20.4	65.8	704	A1226051
9	20	64.5	206	BI121740
10	20	64.5	486	AZ521201
11	20	64.5	491	BI921507
12	20	64.5	589	AQ378111
13	20	64.5	661	BI935904
14	20	64.5	785	CNS02FWO
15	20	64.5	899	BO719125
16	20	64.5	961	BO720141
17	20	64.5	1463	BM044630
18	19.8	63.9	192	AV165009
19	19.8	63.9	310	AQ198424
20	19.8	63.9	338	AQ420157
21	19.8	63.9	485	BM682898
22	19.8	63.9	486	BM719600
23	19.8	63.9	601	BO638029
24	19.8	63.9	641	BO187619
25	19.6	63.2	757	AL602876
26	19.6	63.2	882	CNS0481D
27	19.6	63.2	3014	AK014843
28	19.4	62.6	365	BO782041
29	19.4	62.6	470	BM138265
30	19.4	62.6	484	BE101831
31	19.4	62.6	529	BO162047
32	19.4	62.6	605	BE375801
33	19.4	62.6	607	AZ238089
34	19.4	62.6	670	AZ686355
35	19.4	62.6	692	BO202621
36	19.4	62.6	709	BF572068
37	19.2	61.9	250	BM013755
38	19.2	61.9	340	H81070
39	19.2	61.9	521	TA322C06Q
40	19.2	61.9	523	BI205561
41	19.2	61.9	524	BI205055
42	19.2	61.9	539	TA349604Q
43	19.2	61.9	562	AZ216857
44	19.2	61.9	571	AQ947577
45	19.2	61.9	578	AQ773452

ALIGNMENTS

RESULT 1
LOCUS AJ474604
DEFINITION AJ474604 S00008 Hordeum vulgare cDNA clone S0000800117G07FL, mRNA
ACCESSION AJ474604
VERSION AJ474604.1 GI:21190560
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
REFERENCE 1 (bases 1 to 420)
AUTHORS Saren,A.-M., Tanskannen,J., Paulin,L. and Schulman,A.H.
TITLE Barley EST's
JOURNAL Unpublished (2002)
COMMENT Institute of Biotechnology
Contact: Schulman AH
University of Helsinki
P.O.Box 56 (Vilkinlahti 6A), University of Helsinki FIN-00014, Finland.

FEATURES
source location/Qualifiers
1..420
/organism="Hordeum vulgare"
/db_xref="taxon:4513"
/clone="S0000800117G07FL"
/clone_lib="S00008"
/tissue_type="Callus"

BASE COUNT	/note="Callus K19"
ORIGIN	109 a 103 c 111 g 97 t
Query Match	74.2%; Score 23; DB 9; Length 420;
Best Local Similarity	83.9%; Pred. No. 16;
Matches	26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Oy	1 GCGGAATTCTCAGATGAGGGTCGATTGGCC 31
Dd	255 GCAGAATTTCACATGAGCGTGTTCGCC 285
RESULT 2	
AVB835621/c	574 bp mRNA linear EST 09-MAY-2007
LOCUS	AVB835621 K. Sato unpublished cDNA library: Hordeum vulgare subsp.
DEFINITION	spontaneum top three leaves adult, heading stage Hordeum vulgare subsp. spontaneum cDNA clone bah3110, mRNA sequence.
ACCESSION	AVB835621
VERSION	AVB835621.1 GI:14527710
SOURCE	EST.
ORGANISM	Hordeum vulgare subsp. spontaneum. Hordeum vulgare subsp. spontaneum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; 1 Triticeae; Hordeum. 1 (bases 1 to 574) Sato, K.
REFERENCE	Barley EST sequencing project in NIG and Okayama Univ
AUTHORS	Unpublished (2001)
TITLE	Contact: Kazuhiko Sato
JOURNAL	Research Institute for Bioresources
COMMENT	Okayama University, Barley Germplasm Center Email: kazsato@rib.okayama-u.ac.jp, URL:http://www.rib.okayama-u.ac.jp/barley/ database:http://www.shigen.nig.ac.jp/barley/html. Location/Qualifiers
FEATURES	location=qualifiers
source	1..574 /organism="Hordeum vulgare subsp. spontaneum" /cultivar="H602" /db_xref="taxon:77009" /clone="bah3110" /clone_1lb="K. Sato unpublished cDNA library: Hordeum vulgare subsp. spontaneum top three leaves adult, heading stage" /tissue_type="top three leaves" /dev_stage="adult, heading stage" 138 a 150 c 138 g 146 t 2 others
CASE COUNT	138 a 150 c 138 g 146 t 2 others
ORIGIN	
Query Match	74.2%; Score 23; DB 10; Length 574;
Best Local Similarity	83.9%; Pred. No. 17;
Matches	26; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Oy	1 GCGGAATTCTCAGATGAGGGTCGATTGGCC 31
Dd	216 GCAGAATTTCACATGAGCGTGTTCGCC 186
RESULT 3	
BE885242/c	1119 bp mRNA linear EST 20-OCT-2000
LOCUS	BE885242 NIH_MGC_71 Homo sapiens CDNA clone IMAGE:3907851 5',
DEFINITION	mRNA sequence.
ACCESSION	BE885242
VERSION	BE885242.1 GI:10334018
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE	1 (bases 1 to 1119)
AUTHORS	NIH-MGC http://mgc.nhl.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L14M9719 row: a column: 04 High quality sequence stop: 231. Location/Qualifiers
FEATURES	1..1119
source	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3907851" /clone_11b="NIH-MGC_71" /tissue_type="TelomYosarcoma" /lab_host="DH10B (phage-resistant)" /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1; Notif: Site 2: Salt: Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.1 kb."
BASE COUNT	481 a 291 c 305 g 42 t
ORIGIN	
Query Match	74.2%; Score 23; DB 12; Length 1119;
Best Local Similarity	100.0%; Pred. No. 19;
Matches	23; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
QY	9 CTCACATGAGGGGTGCATTGGCC 31
Db	206 CTCACATGAGGGGTGCATTGGCC 184
LOCUS	
DEFINITION	ACQ349812 757 bp DNA linear GSS 07-MAY-1999
ACCESSION	ACQ349812
VERSION	ACQ349812.1 GI:4177147
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 757) Zhao, S., Adams, M.D., Niernan, W., Malek, J., de Jong, P. and Venter, J.C.
TITLE	Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
JOURNAL	Unpublished (1997)
COMMENT	Other-GSSs: RPC11-118N20_TV Contact: Shaying Zhao, William Niernan, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbeet@igr.org Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pietere@jeng.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.igr.org/tadb/humgen/bac_end_search/bac_end_search.html Seq primer: Sp6 Class: BAC ends. Location/Qualifiers
FEATURES	

```

SOURCE
1. .757
/organism="Homo sapiens"
/db_xref="GDB:7545259"
/db_xref="taxon:9606"
/clone="RPC1-11-118N20"
/clone_id="RPC1-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC library"
BASE COUNT      206 a      166 c      137 g      246 t      2 others
ORIGIN
Query Match      68.4%; Score 21.2; DB 17; Length 757;
Best Local Similarity 88.5%; Pred. No. 1.1e+02;
Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCGGAATTCACATGAGGTCGATC 26
|||||
Db 6 GCGGAATTCACATGAGGTCGATC 31

RESULT 5
BF796120/c      984 bp      mRNA      linear      EST 12-JAN-2001
LOCUS           602258969F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:4342360 5',
DEFINITION      mRNA sequence.
ACCESSION       BF796120
VERSION         BF796120.1 GI:12101174
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE       1 (bases 1 to 984)
AUTHORS         NIH-MGC http://mgc.nci.nih.gov/.
TITLE           National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL         Unpublished (1999)
COMMENT         Contact: Robert Strausberg, Ph.D.
                  Email: cgabs@mail.nih.gov
                  Tissue Procurement: Louis Staudt, M.D., Ph.D.
                  cDNA Library Preparation: Life Technologies, Inc.
                  DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
                  DNA Sequencing by: Incyte Genomics, Inc.
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LNLN at:
                  http://image.llnl.gov
                  Plate: LHAM9957 row: 1 column: 17
                  High quality sequence start: 13
                  High quality sequence stop: 628.
FEATURES
source
1. .984
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="NIH_MGC_85"
/tissue_type="Lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."
BASE COUNT      248 a      263 c      290 g      183 t
ORIGIN
Query Match      67.7%; Score 21; DB 12; Length 984;
Best Local Similarity 82.8%; Pred. No. 1.4e+02;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 GGAATTCACATGAGGTCGATTTGCC 31
|||||
Db 870 GGTATTCACATGAGGTCGATTTCCC 842

```

```

RESULT 6
BG955224
LOCUS           CM4-CT0657-120201-861-910 CT0657 Homo sapiens cDNA, mRNA sequence.
DEFINITION      BG955224
ACCESSION       BG955224
VERSION         BG955224.1 GI:14373395
KEYWORDS        EST.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE       1 (bases 1 to 385)
AUTHORS         Dias Neto,E., Garcia Correa,R., Verjovsky-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W., Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE           Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL         Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE         20202663
COMMENT         Contact: Simpson A.J.G.
                  Laboratory of Cancer Genetics
                  Ludwig Institute for Cancer Research
                  Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
                  Brazil
                  Tel.: +55-11-2704922
                  Fax: +55-11-2707001
                  Email: asimpson@ludwig.org.br
                  This sequence was derived from the FAPESP/LICR Human Cancer Genome
                  Project. This entry can be seen in the following URL
                  (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM4ct2-CM4-CT0657-
                  120201-861-910&t3=2001-02-12&t4=1)
                  Seq primer: puc 18 forward
                  High quality sequence start: 69
                  High quality sequence stop: 385.
FEATURES
source
1. .385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="CT0657"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT      97 a      55 c      68 g      165 t
ORIGIN
Query Match      66.5%; Score 20.6; DB 13; Length 385;
Best Local Similarity 85.2%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GAATTCACATGAGGTCGATTTGCC 30
|||||
Db 220 GAATTCACATGAGGTCGATTTCCC 246

RESULT 7
AM128188/c
LOCUS           AM128188      528 bp      mRNA      linear      EST 25-OCT-1999
DEFINITION      f111f07.x1 Sugano Kawakami zebrafish DRF Danio rerio cDNA clone
2600869 3' similar to contains element MER30 repetitive element ;,
mRNA sequence.
ACCESSION       AM128188
VERSION         AM128188.1 GI:6116122
KEYWORDS        EST.

```

SOURCE	zebrafish.
ORGANISM	Danio rerio.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes; Cyprinidae; Danio.
AUTHORS	1 (bases 1 to 528) Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S., Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T., Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter,E., Bowers,T., Wylie,T., Waterson,R. and Wilson,R.
TITLE	Washu Zebrafish EST Project 1999
JOURNAL	Unpublished (1999)
COMMENT	Contact: S.L. Johnson Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami Sequencing by: Washington University Genome Sequencing Center Seq primer: T7 ET from Amersham High quality sequence stop: 479.
FEATURES	Location/Qualifiers 1..528 /organism="Danio rerio" /strain="AB" /db_xref="taxon:7955" /clone="2600869" /clone_1lb="Sugano Kawakami zebrafish DNA" /sex="mixed (one male and one female, including unfertilized eggs)" /dev_stage="adult" /lab_host="DH10B (phage resistant)" /note="Vector: pME185-FL3; Site_1: DraIII (CAGCTGTGTG); Site_2: DraIII (CAGCATGTGT); 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCCCTTTTCTTTTCTTTTCTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TCTTGGCCCTACTGTG], digested and cloned into distinct DraIII sites of the pME185-FL3 vector (5' site CAGCTGTG, 3' site CAGCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' end primer CTTCGCTCTAAAGACTGTG and 3' end primer CGACCTGCAGCTCGACACA."
BASE COUNT	193 a 92 c 74 g 169 t
ORIGIN	
Query Match	66.5%; Score 20.6; DB 10; Length 528;
Best Local Similarity	85.2%; Pred. No. 1.8e+02;
Matches	23: Conservative 0; Mismatches 4; Indels 0; Gaps 0;
OY	3 GGAATTCACATGAGGCGCATTTGC 29
Db	503 GGAATTCACATGAGGCGCATTTTC 477
RESULT 8	
LOCUS	A1226051
DEFINITION	U108f10.y1 Sugano mouse liver m1a Mus musculus cDNA clone IMAGE1891335 5' similar to gb:J04973 UBIQUINOL-CYTTOCHROME C REDUCTASE CORE PROTEIN 2 PRECURSOR (HUMAN); mRNA sequence.
ACCESSION	A1226051
VERSION	A1226051.1 GI:3809104
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 704) Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
AUTHORS	

TITLE	The WashU-HHMI Mouse EST Project						
JOURNAL	Unpublished (1996)						
COMMENT	Contact: Maria M/Mouse Est Project WashU-HHMI Mouse EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:975663						
FEATURES	Trace considered overall poor quality Seq primer: custom primer used High quality sequence stop: 1.						
SOURCE	Location/Qualifiers 1..704 /organism="Mus musculus" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:1891339" /clone_id="Sugano mouse liver mlia" /sex="female" /dev_stage="adult" /lab_host="DH10B" /note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII (CACCTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGGGCCCTTTTCTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor (TGTGGCCACTGTG), digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCGTGCTCTAAAGCTGCG and 3' end primer CGACCTGCAGCTGCAGCAC."						
BASE COUNT	190 a 158 c 172 g 184 t						
ORIGIN							
Query Match	65.8%; Score 20.4; DB 9; Length 704;						
Best Local Similarity	95.5%; Pred. No. 2.3e+02;						
Matches 21:	Conservative	0;	Mismatches	1;	Indels	0; Gaps	0;
OY	5 AATTCTCACATGAGGTCATT 26 Db 156 AATTCTCACATGAGGTCATT 177						
RESULT 9	Bi121740 206 bp mRNA linear EST 31-DEC-2001						
LOCUS	P045P75J Populus flower cDNA library Populus balsamifera subsp.						
DEFINITION	trichocarpa cDNA, mRNA sequence.						
ACCESSION	Bi121740						
VERSION	Bi121740.1 GI:18005715						
KEYWORDS	EST.						
SOURCE	Populus balsamifera subsp. trichocarpa.						
ORGANISM	Populus balsamifera subsp. trichocarpa Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids I; Malpighiales; Salicaceae; Populus. 1 (bases 1 to 206) Herzberg,M., Aspeborg,H., Gustafsson,R., Bjorkbacka,H., Hiltunen,T., Karlsson,J., Teeril,T., Ekstrandson,P., Bahlerna,R., Jansson,S., Nilsson,O., Sundberg,B., Nilsson,P., Uhlen,M., Sandberg,G. and Lundberg,J.						
REFERENCE	Gene expression in Populus						
AUTHORS	Lundberg,J.						
TITLE	Gene expression in Populus						

ACCESSION AO378111
 VERSION AO378111.1 GI:4349134
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 589)
 AUTHORS Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.
 TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building
 JOURNAL Unpublished (1997)
 COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: hbeetlgr.org
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (inforesgen.com). BAC end search page: http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
 Seq primer: SP6
 Class: BAC ends.
 FEATURES
 source Location/Qualifiers
 1..589
 /organism="Homo sapiens"
 /db_xref="GDB:7561886"
 /db_xref="taxon:9606"
 /clone="RPCI-11-162C15"
 /clone_lib="RPCI-11"
 /sex="Male"
 /cell_type="Lymphocytes"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male BAC Library"
 BASE COUNT 177 a 121 c 137 g 153 t 1 others
 ORIGIN
 Query Match 64.5%; Score 20; DB 17; Length 589;
 Best Local Similarity 82.1%; Pred. No. 3.3e+02;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Oy 1 GCGGAAATTCACATGAGGTCATTG 28
 ||||||| ||||||| ||| |||
 b 169 GCGGAAATTCACATGAGGACAGATTG 196
 ||||||| ||||||| ||| |||
 RESULT 13
 BI935904/c 661 bp mRNA linear EST 18-OCT-2001
 LOCUS EST555793 tomato flower, anthesis Lycopersicon esculentum cDNA
 DEFINITION clone cTOD24124 5' end, mRNA sequence.
 ACCESSION BI935904
 VERSION BI935904
 KEYWORDS
 SOURCE
 ORGANISM
 Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Lycopersicon.
 TITLE 1 (bases 1 to 661)
 AUTHORS van der Hoeven, R.S., Bezzerides, J.L., Karamycheva, S.A., Tsai, J., Utechtack, T., Van Aken, S., Rinning, C.M., Nierman, W., Fraser, C.M., Martin, G.B., Giovannoni, J.J. and Tanksley, S.D.
 COMMENT Generation of ESTs from tomato flower tissue, anthesis (2001)
 JOURNAL Unpublished (2001)
 CONTACT: CUGI
 CLEMSON UNIVERSITY GENOMICS INSTITUTE

CLEMSON UNIVERSITY
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: http://www.genome.clemson.edu/orders/index.html
 This clone is available through the Clemson University Genomics Institute
 Seq primer: T3.
 FEATURES
 source Location/Qualifiers
 1..661
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="cTOD24124"
 /clone_lib="tomato flower, anthesis"
 /tissue_type="flower"
 /dev_stage="anthesis"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; supplier: Cornell University; sequencing: The Institute for Genomic Research; flower buds and flowers were taken from greenhouse plants (4-8 wks old, TA496). They were immediately frozen in liquid nitrogen and then size-separated while remaining frozen."
 BASE COUNT 183 a 92 c 120 g 266 t
 ORIGIN
 Query Match 64.5%; Score 20; DB 13; Length 661;
 Best Local Similarity 82.1%; Pred. No. 3.4e+02;
 Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 Oy 4 GAATTCACATGAGGTCATTGCC 31
 ||| ||||||| ||| |||
 Db 281 GAATTCACATGAGGTCATTCCCC 254
 ||| ||||||| ||| |||
 RESULT 14
 BI935904/c 785 bp DNA linear GSS 13-MAY-2000
 LOCUS CNS02FVQ
 DEFINITION Tetradon nigroviridis genome survey sequence PUC-Orl end of clone 135M02 of library G from Tetradon nigroviridis, genomic survey sequence.
 ACCESSION AL195551.1 GI:7833657
 VERSION AL195551
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetradon nigroviridis.
 ORGANISM Tetradon nigroviridis.
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
 REFERENCE 1 (bases 1 to 785)
 AUTHORS Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fitzames, C., Winkler, P., Brotlier, P., Quetier, F., Saurin, W. and Weissenbach, J.
 COMMENT Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence
 UNPUBLISHED
 2 (bases 1 to 785)
 REFERENCE Roest-Crollius, H., Jallion, O., Dasilva, C., Fitzames, C., Fisher, C., Bouneau, L., Billaud, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 785)
 AUTHORS
 TITLE
 JOURNAL
 COMMENT Direct Submission
 Submitted (12-APR-2000)
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetradon.
 FEATURES
 source Location/Qualifiers
 1..785
 /organism="Tetradon nigroviridis"

/db_xref="taxon:99883"
/clone="135M02"
/clone_1lb="G"
/note="Genoscope sequence ID : COAG135B01SP1-end :
PUC-ori"

BASE COUNT 197 a 185 c 169 g 176 t 58 others

ORIGIN

Query Match 64.5%; Score 20; DB 17; Length 785;
Best Local Similarity 82.1%; Pred. No. 3.5e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 GAATTCACATGAGGTGCATTGCCC 31
|||||
Db 207 GCATGCTCACATGAAGTTCATGTGCCC 180

RESULT 15
BQ719125 899 bp mRNA linear EST 16-JUL-2002
LOCUS
DEFINITION AGENCOURT_8232632_lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6191580 5', mRNA sequence.
ACCESSION BQ719125
VERSION BQ719125.1 GI:21858022
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 899)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. James R. Lupski
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM13592 row: d column: 13
High quality sequence stop: 678.
Location/Qualifiers
1. 899

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6191580"
/clone_1lb="lupski_sympathetic_trunk"
/sex="male"
/tissue.type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCAGCGTCGC-3' and
5'-GACTACTTCAGATCGGAGCGGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

BASE COUNT 230 a 176 c 249 g 244 t

ORIGIN

Query Match 64.5%; Score 20; DB 14; Length 899;
Best Local Similarity 82.1%; Pred. No. 3.6e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 GAATTCACATGAGGTGCATTGCCC 31
|||||

Db 346 GAATTCACACAGGAGTGCATTGTCC 373
Search completed: March 23, 2003, 17:12:25
Job time : 206.545 secs

THIS PAGE BLANK (USPTO)